Patient Preference and Adherence

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Patient Preferences in Infertility Treatment in China

Jiali Chen[1], Xiaodan Qian[2], Dan Su[2], Jinhong Gong[2], Jingjing Shang[2], Lingli Zhang[3], Xin Li[1,2,4,5]

[1]Department of Health Policy, School of Health Policy and Management, Nanjing Medical University, Nanjing, People's Republic of China; [2]Department of Pharmacy, The Second People's Hospital of Changzhou, The Third Affiliated Hospital of Nanjing Medical University, Changzhou, People's Republic of China; [3]School of International Pharmaceutical Business, China Pharmaceutical University, Nanjing, People's Republic of China; [4]Department of Pharmaceutical Regulatory Science and Pharmacoeconomics, School of Pharmacy, Nanjing Medical University, Nanjing, People's Republic of China; [5]Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, People's Republic of China

Correspondence: Xin Li, Nanjing Medical University, Nanjing, 211166, People's Republic of China, Email xinli@njmu.edu.cn; Lingli Zhang, China Pharmaceutical University, Nanjing, 211198, People's Republic of China, Email zhanglingli@cpu.edu.cn

Background: Infertility has become a growing public health problem in China. It is important to accurately and easily obtain patient preferences. This study aimed to obtain and compare stated preference results from the perspective of infertility patients. By assessing the validity and acceptability of both methods, it provides lessons for empirical research and practical application.

Methods: Discrete choice experiments (DCE) and profile case best-worst scaling (BWS-2) are methods used to identify and weigh various criteria affecting the order of priorities. We conducted a face-to-face survey of female patients with infertility aged between 20–45 years. The survey included socio-demographic information, preference questionnaires and completion of evaluation questions. Attributes included live birth rate, pregnancy rate, degree of participation in treatment decision making, maternal complications, neonatal complications and program cost. Conditional logit models were used to analyze attribute level weights and relative importance was calculated separately.

Results: A total of 330 valid questionnaires were collected. The preferences of patients experiencing infertility were quantified through two stated preference research methods. The findings indicated that patients exhibited a preference for treatment options that were highly effective, exhibited minimal side effects, were patient-centered, and were cost-effective. The BWS-2 and DCE preference weights demonstrated high consistency, with only slight difference observed in the ranking of individual attributes within the order of relative importance. In the view of the patients, the DCE questions were perceived to be less challenging to comprehend and were therefore preferred to be completed.

Conclusion: The BWS-2 and DCE exhibit identical validity and highly consistent preference results. In the context of specific research questions, the selection of a method or the combination of methods must be suitable to the purpose of the study in order to ensure that the utility gained is maximized. Further research is required to corroborate these findings.

Keywords: discrete choice experiment, best-worst scaling, infertility treatment, preferences

Introduction

Preference can be defined as a specific expression of the maximum value of utility.[1] In this context, treatment willingness and choice preference can be understood as rational choices made by patients after weighing different treatment situations. These choices reflect the individual's assessment of the value of health services and prioritization.[2] An understanding of patient preferences is meaningful in the development of a patient-centered healthcare service model,[3] which is a crucial element in the provision of high-quality healthcare services and an effective means of enhancing the public's healthcare experience and contributing to the advancement of the healthcare industry. Among the stated

Received: 30 October 2024
Accepted: 22 March 2025
Published: 30 March 2025

preference research methods, discrete choice experiments (DCE) and profile case best-worst scaling (BWS-2) are two typical measurement tools. Both National institute for Health and Care Excellence and Food and Drug Administration recommend the application of such methods to obtain patients' preference information.[4] However, there are notable differences in experimental design, analytical models, and participants' behavioral decisions.[5–7] In details, they differ in the approach to choice set design. BWS requires patients choose the best and worst options at the intrinsic attributes of the object, frequently employing counting and modelling methods for data analysis. DCE requires patients to make choices in a virtual situation consisting of levels of attributes (trade-offs), often using modelling methods for data analysis. In summary, BWS-2 captures the advantages of patient prioritization through extreme value selection. In contrast, DCE is effective in modelling complex clinical decision-making scenarios. In the field of medicine and healthcare, the ISPOR organization has presented best practices and guidelines for DCE preference studies.[8–10] While BWS-2 was introduced to healthcare research relatively recently and has relevant guidelines for its use,[11] it has unique advantages as an emerging research method.[12] The advantages of BWS are demonstrated by the decreasing sample size and number of attributes in the study, as well as the relative simplicity of the analysis method. BWS is an attractive and relatively easy method. It is therefore important to compare the preferred outcomes of the two methods and to assess patients in order to promote the application of both methods.

In recent years, China's fertility and population growth rates have continued to decline, and the incidence of infertility has continued to climb, which has become a serious problem affecting social development.[13] The prevalence of infertility in China has increased from 11.9% in 2007 to 17.6% in 2020. The infertility rate among Chinese women is estimated to be as high as 15%-20%.[14] A quarter of couples of reproductive age suffer from infertility, compared to the global average of 15%.[15] This contributes significantly to an ageing population and a shrinking labor force. For patients, it is a heavy financial burden and a social stigma.[16] Infertility patients tend to choose assisted reproductive technologies (ART) to achieve their fertility wishes.[17] Techniques such as embryo transfer and preimplantation genetic screening have increased the probability of patients having children to varying degrees.[18] However, there are significant differences in specific procedures, treatment side effects, and maternal and infant safety, all of which influence patients' choice preferences.[19] The diverse applicability of multiple techniques makes treatment decisions highly preference sensitive.[20] Relevant studies emphasize the significance of patient preferences for decision-making or policy implementation in the field of assisted reproduction. For example, Harrison et al[21] found that patients' willingness correlates with stronger perceived control and proactive decision-making. This emphasizes the need for clinics to adopt a multi-cycle framework during consultations that aligns with patient preferences. In addition, patients value cost during treatment, particularly embryo donation in frozen embryo transfer,[22] which complicates decision-making. In DCE study, patients in the Netherlands realized two patterns of preference, valuing success or valuing low-burden.[23] Therefore, it is important to accurately and efficiently measure patient preferences and incorporate them into multi-cycle plans using a shared decision-making model that balances efficacy, burden and other factors. In China fertility treatments are gradually being included in health insurance payments and patient out-of-pocket costs are decreasing. In addition, the adoption of patient-centered approaches, such as providing more information about treatment and helping patients to participate in decision-making, may also alter patient choice.[24] ARTs are rapidly evolving and new treatment options are complicating decision-making.[25] The selection of safe, cost-effective, and appropriate treatment options that incorporate patient preferences has become key to improving the quality of care and facilitating scientific decision-making,[26] which can help to optimize the level of healthcare services, improve patient compliance as well as improve pregnancy outcomes.

A number of international studies have been conducted on the applicability of preference measurement tools, in areas such as disease treatment and prioritization. However, there is an absence of research related to the preference measurement tools in infertile patients. The objective of this study was to identify the factors that influence the treatment preferences of infertility patients and to examine the differences in the preferred outcomes of patients when surveyed using the BWS-2 and DCE questionnaire formats. Furthermore, patients' assessments of the two methods were compared in order to provide lessons and references for empirical research and practical use of disease treatment preferences.

Method

Study Population

Currently, ART has been implemented on a significant scale, with 622 medical institutions in China having obtained approval to conduct human assisted reproduction technology. Of these, 38 are located in Jiangsu Province, which is the second largest region in China. The inclusion criteria for patients were: women aged 20–45 years, suffering from infertility, seeking medical assistance for infertility, having good communication skills, and voluntary participation in the questionnaire study. The study was conducted in three hospitals, of which Nanjing Drum Tower Hospital and Jiangsu Province Hospital both have hospital districts exclusively dedicated to carrying out assisted human reproduction technology, while the research in Jiangsu Women and Children Health Hospital was carried out in the reproductive medicine center. The study was conducted in the female-specific clinic rooms of the three sample hospitals. Participants were recruited by simple random sampling from patients who met the criteria. At present, there is a paucity of uniform international standards for the calculation of sample sizes in BWS-2 studies. The most common method employed is to draw upon DCE to determine the requisite sample size. A minimum of 84 respondents (up to three levels, nine choice sets and two alternatives) were required, as determined according to the Orme's formula ($N > 500*c/(t*a)$).[27] Further details are available in the eAppendix 1. In consideration of the response rate and tolerable error, and given that the objective of the study was not to account for heterogeneity of preferences, a sample size of 300 patients was planned to ensure reliable statistical analyses. The anonymous research was conducted between April and June 2024. Prior to the administration of the questionnaire, respondents were provided with comprehensive information regarding the purpose and content of the survey by a trained researcher. The study was approved by the Ethics Committee of Nanjing Medical University, with ethics approval number 2020103. All methods were performed in accordance with the relevant guidelines and regulations outlined in the Declaration of Helsinki. Informed consent was obtained from all participants.

Attributes and Levels

This study was based on an extensive literature review, which identified three main categories of attributes:[28] outcome attributes, process attributes, and cost attributes. In total, 13 attributes were identified (details in eTable 1). An object case BWS-1 was employed to prioritize the attributes from the patient's perspective. The study was conducted with 150 patients. The results of this study (see in eTable 2) led to the finalization of six attributes, as detailed in Table 1. To ensure that the study design was practical and feasible within the current healthcare and policy frameworks, an expert consultation was conducted with an associate professor of medicine and a nurse from a reproduction center. This consultation was used to further refine and modify the formulation of the attributes. Following this, the formulation of the identified attributes was further refined and modified. Refinements and modifications were then made to the questionnaires designed for the identified attributes, and a pilot study was conducted. A sample of 40 patients was selected from Nanjing Drum Tower Hospital to assess the practicality, acceptability, and reliability of the study. The final questionnaire comprised three sections. The first section collected basic information about the patients, while the second section consisted of a series of choice sets. The order of the questions in the DCE and the profile case BWS-2 was randomized to avoid any potential bias due to differences in the order of completion.[29] Patients were asked to choose their preferred option after a detailed and concise explanation of the questionnaire's content, structure, and requirements for completion. Ultimately, it is essential to ascertain the patients' assessment of the two preference measurement tools.

BWS-2 Questionnaire Design

In this study, 18 distinct profile case BWS-2 choice sets, each comprising six attributes and one of their corresponding levels as previously described, were constructed using orthogonal main effect design (OMED) using R 4.3.3 software. To minimize the burden on patients, the questionnaires were distributed equally between the two versions. Furthermore, each version was supplemented with a single set of consistency test questions. In order to facilitate comprehension of the questions posed in this study, attribute and level explanation cards and corresponding sample questions were prepared. See Figure 1 for details.

Table 1 Attributes and Associated Levels Identified for This Stated Preference Study

Attributes	Levels	Description
Live birth rate	15% 35% 55%	The rate of live births is the main measure of treatment success. For example, a live birth rate of 15% means that 15 out of 100 patients deliver successfully.
Clinical pregnancy rate	20% 40% 60%	Refers to the patient's pregnancy rate per treatment cycle. For example, a clinical pregnancy rate of 20% means that 20 out of 100 patients become pregnant.
Degree of participation in treatment decision making (TDM)	Not participate in TDM Partially participate in TDM Fully participate in TDM	Refers to the timely communication of the patient's views and concerns to the physician and the participation in the discussion and selection of treatment options. a) The patient is not involved in the treatment decision making, and the physician makes all treatment decisions based entirely on his or her own expertise, without input from the patient. b) The patient is partially involved in the treatment decision making, and the physician explains some but not all of the treatment options to the patient. c) The patient is fully involved in the treatment decision making, and the physician explains all treatment options to the patient and supports the patient in making the best treatment decision for him or her.
Risk of maternal complications	0.1% 2.5% 5%	Refers to the likelihood of maternal complications such as moderate to severe ovarian hyperstimulation syndrome (OHSS), postoperative bleeding and postoperative pelvic infection. The most serious of these complications is OHSS, which can lead to a range of potential symptoms from moderate abdominal swelling to more severe swelling and abdominal pain to extreme thirst and dyspnea.
Risk of neonatal complications	5% 10% 15%	Refers to the likelihood of complications such as prematurity (before 37 weeks), low birth weight and congenital anomalies (birth defects) in newborns. These complications may result in the infant being admitted to the neonatal intensive care unit for a period of time and in some cases may lead to longer periods of varying severity of health problems.
Out-of-pocket costs per cycle (¥)	3000 30,000 50,000	Refers to the out-of-pocket expenses of the patient for the treatment. Assisted reproduction in China is not covered by basic health insurance and is entirely at the individual's expense. [a]

Notes: [a] During the survey, assisted reproductive technologies were not covered by health insurance.

DCE Questionnaire Design

Similarly, 18 choice sets were generated and randomized into two versions using bayesian efficient design using Ngene 1.4 software. The cond command was used in the design to avoid the implausible situation where the live birth rate is higher than the pregnancy rate. In addition, patients were given the option to decline healthcare services regardless of their attributes and level of care. The unforced choice provided an opt-out option that required respondents to complete the forced choice and choose between opt-out options.[30,31] Each version was supplemented with an additional set of consistency testing questions. The questionnaire design for this study was conducted in accordance with the ten criteria for stated preference research set forth by the ISPOR.

Statistical Analysis

Both BWS-2 and DCE are founded on random utility theory, which is employed for the purpose of eliciting patients' stated preferences. The BWS-2 data was transformed with the marginal sequential model using R 4.3.3 software. The counting method was calculated based on the frequency with which the attribute levels were selected as the best and worst. The attribute level with the lowest utility in the count model was employed as a reference term for further conditional logit model analysis. Significant positive and negative coefficients indicate a preference for certain attribute levels. The same conditional

Best		Worst
o	Live birth rate: 35%	o
o	Clinical pregnancy rate: 60%	o
o	Degree of participation in treatment decision making: full	o
o	Risk of maternal complications: 10%	o
o	Risk of neonatal complications: 5%	o
o	Out-of-pocket costs per cycle: ¥3000	o

	Option A	Option B
• Live birth rate	15%	35%
• Clinical pregnancy rate	60%	40%
• Degree of participation in treatment decision making	No participation	Partial participation
• Risk of maternal complications	5%	5%
• Risk of neonatal complications	10%	15%
• Out-of-pocket costs per cycle (¥)	30,000	50,000
Which treatment option do you prefer?	☐	☐
In real life, would you accept the above program?	☐ accept ☐ not accept	

Figure 1 Example of BWS-2 and DCE choice tasks.

logit model analysis was carried out for the DCE data using Stata 17 software. In order to facilitate comparison between DCE and BWS-2, the relative importance of attributes was calculated separately. A comparative analysis of the two preference measurement tools was conducted by plotting the scatterplot of the main effects model and measuring the correlation of the preference results under the different preference tools. Additionally, the results of the evaluation questions of the two methods were analyzed. The level of statistical significance was determined using a threshold of $P<0.05$ (two-sided), and 95% confidence intervals (CI) were calculated using the delta method.

Results

A total of 40 patients diagnosed with infertility completed the questionnaire utilized for the prior survey. The model estimates demonstrated positive and negative signs and weights (see eTable 3). This indicates that respondents are able to comprehend and complete the DCE choice sets. The prior values and patient feedback will be applied to the DCE choice set design, with modifications to the final version of the formal questionnaire.

Basic Information

A total of 400 respondents were invited to complete the questionnaire. A total of 330 valid questionnaires were included in the study, with a high-quality control pass rate, following the exclusion of sample that did not complete all questions and did not pass the consistency test. The results are demonstrated in Table 2. The mean age of the respondents was 32.9 years old. Only 23.6% of them had their usual place of residence in the city. They were well-educated, with 64.6% having obtained a university or college degree or above. Additionally, half of them had a monthly household income within the range of $743 and $2,226. With regard to the patients' experiences of treatment, 44.8% had been preparing for pregnancy for 3–5 years, over half of the patients had received previous fertility treatment, and 65.8% were currently engaged in the treatment cycle.

https://doi.org/10.2147/PPA.S501723

Table 2 Characteristics of Respondents Included for Analysis

	Frequency (N=330)	Percentage (%)
Age (years)		
20–29	72	21.8
30–34	149	45.2
35–39	79	23.9
40–45	30	9.1
Residence		
Within the city	78	23.6
Outside the city	126	38.2
Outside the province	126	38.2
Education		
Junior high school and below	43	13.0
High school or junior college	74	22.4
College or undergraduate	186	56.4
Master or doctor	27	8.2
Work		
Work less than 40 hours	108	32.7
Work more than 40 hours	111	33.6
No work	111	33.6
Average monthly household income		
Less than ¥5000	60	18.2
Between ¥5000 and ¥15000	193	58.5
More than ¥20000	77	23.3
Pregnancy preparation time		
Less than 2 years	130	39.4
Between 3 and 5 years	148	44.8
Between 6 and 9 years	39	11.8
More than 10 years	13	3.9
Received treatment in the past		
Yes	222	67.3
No	108	32.7
Current stage of treatment		
First visit	33	10.0
Physical examination	72	21.8
Accepting ART	217	65.8
Fetal preservation	8	2.4
Factors leading to infertility		
Male	32	9.7
Female	160	48.5
Both parties	56	17.0
Unknown	82	24.8
Self-reported health status		
Very good	51	15.5
Good	112	33.9
Average	163	49.4
Bad	3	9
Very bad	1	3

Comparison of Preferences

The results of the model in the BWS-2 (Table 3) show that using neonatal complications as the reference attribute, the magnitude of the regression coefficient value indicates the magnitude of the utility value obtained at that level in comparison to the reference level. The conditional logit model results showed the highest positive impact was the live birth rate attribute

Table 3 Results of the Profile Case best-Worst Scaling (BWS-2) Analysis

	Counting Model			Conditional Logit Model		
	Best	**Worst**	**Std.BW[a]**	**Coef.**	**SE**	**p-value**
Attribute impacts						
Live birth rate				2.58	0.07	0.00
Clinical pregnancy rate				2.79	0.08	0.00
Participation in TDM				1.86	0.08	0.00
Maternal complications				Ref.		
Neonatal complications				1.34	0.07	0.00
Costs per cycle (¥)				0.79	0.06	0.00
Level scale values						
Live birth rate						
15%	27	120	−0.09	Ref.		
35%	199	18	0.17	0.12	0.08	0.14
55%	370	10	0.33	1.91	0.08	0.00
Clinical pregnancy rate						
20%	105	81	0.02	Ref.		
40%	202	13	0.17	−0.09	0.08	0.26
60%	312	3	0.28	1.37	0.08	0.00
Participation in TDM						
No participation	66	87	−0.02	Ref.		
Partial participation	43	36	0.01	0.02	0.09	0.85
Full participation	78	16	0.06	0.96	0.10	0.00
Maternal complications						
5%	97	28	0.06	Ref.		
10%	43	67	−0.02	−0.10	0.09	0.23
15%	21	119	−0.09	−0.71	0.09	0.00
Neonatal complications						
0.1%	11	181	−0.16	Ref.		
2.5%	3	273	−0.25	−0.03	0.08	0.71
5%	3	246	−0.22	−0.46	0.07	0.00
Costs per cycle						
¥3000	36	20	0.01	Ref.		
¥30,000	9	107	−0.09	−0.23	0.08	0.01
¥50,000	4	203	−0.18	−0.83	0.08	0.00
Likelihood ratio test=3155						

Notes: [a] Std.BW: The value is zero when the attribute is selected as the best as many times as the worst. The score ranges from −1 to 1.

at the 55% level (1.91, p<0.001), and the negative impact was the $50,000 project cost (−0.83, p<0.001). Patients had no significant preference for the intermediate levels of each attribute (p>0.1). These findings were also reflected in the priority of Std.BW in the counting method. To directly estimate attributes weights and level preferences, the BWS-2 utilized effect coding. In order to ensure consistency in the coding process, both effect variable coding and dummy variable coding were applied to the data collected for the DCE in this study. Table 4 presents the regression model results for DCE. It can be seen that the preference weights are similar to those of BWS-2, with a monotonically increasing trend in the preference weights within the attributes, with the exception of the two complications attributes, which have low values. An increase in the live birth rate from 15% to 55% would result in the patient gaining a marginal utility of 0.77 (p < 0.001), which is 4.5 times the lowest utility value. Furthermore, the coefficient of the exit option in the DCE study was 0.31, indicating that opting out yielded a positive value utility for the respondents and that infertility patients were more inclined to abandon the treatment.

https://doi.org/10.2147/PPA.S501723

Table 4 Results of the Discrete Choice Experiment (DCE) Analysis

	Effect Coding			Dummy Coding		
	Coef.	**SE**	**p-value**	**Coef.**	**SE**	**p-value**
Live birth rate						
15%	Ref.					
35%	0.14	0.19	0.46	0.59	0.25	0.02
55%	0.32	0.14	0.02	0.77	0.12	0.00
Clinical pregnancy rate						
20%	Ref.					
40%	−0.16	0.11	0.13	0.03	0.16	0.84
60%	0.35	0.07	0.00	0.54	0.07	0.00
Participation in TDM						
No participation	Ref.					
Partial participation	0.06	0.06	0.28	0.30	0.13	0.02
Full participation	0.17	0.04	0.00	0.40	0.11	0.00
Maternal complications						
5%	Ref.					
10%	−0.18	0.09	0.05	−0.40	0.14	0.01
15%	−0.04	0.05	0.43	−0.26	0.07	0.00
Neonatal complications						
0.1%	Ref.					
2.5%	−0.10	0.07	0.15	−0.28	0.11	0.01
5%	−0.08	0.04	0.07	−0.26	0.06	0.00
Costs per cycle						
¥3000	Ref.					
¥30,000	−0.17	0.05	0.00	−0.17	0.05	0.00
¥50,000	−0.29	0.09	0.00	−0.29	0.09	0.00
Option-out	0.31	0.07	0.00	0.31	0.07	0.00
Log-likelihood=−1314.73						
AIC=2653.47						
BIC=2730.09						

The results of the present study demonstrated statistically significant findings for both the DCE and BWS-2 preference studies, indicating that both preference measurement tools exhibited validity in this study. The BWS-2 was observed to be more responsive to alterations in level (from level 1 to level 2). For instance, the estimated relative utility change associated with a shift in the live birth rate from a medium (35%) to a high (55%) level was considerably higher in the BWS-2 than in the DCE. A Spearman correlation analysis was conducted on the data from the two models, and the results demonstrated a strong positive correlation between the preference results of the two methods, with a correlation coefficient of 0.811 ($P < 0.001$). Furthermore, the coefficient of determination of the linear fit in the scatter plot (Figure 2) is 0.76, which also indicates a high level of agreement between the preference results of the two measurement tools. Furthermore, the slope of the fitted line is 3.56, indicating that the difference in coefficient sizes across attribute levels in BWS-2 is considerably larger than in DCE for equal unit changes.

Further analysis reveals difference between the relative importance results of the two methods. In order to facilitate comparison between studies, this study sets the live birth rate attribute with the highest utility value as the common denominator for scaling. As illustrated in Figure 3, the difference between the two method preference outcomes is specifically seen in the importance of program costs, which ranked fourth in BWS-2, much higher than neonatal complications, and fifth in DCE. Maternal complications were also elevated in importance in DCE to almost the same status as participation in TDM.

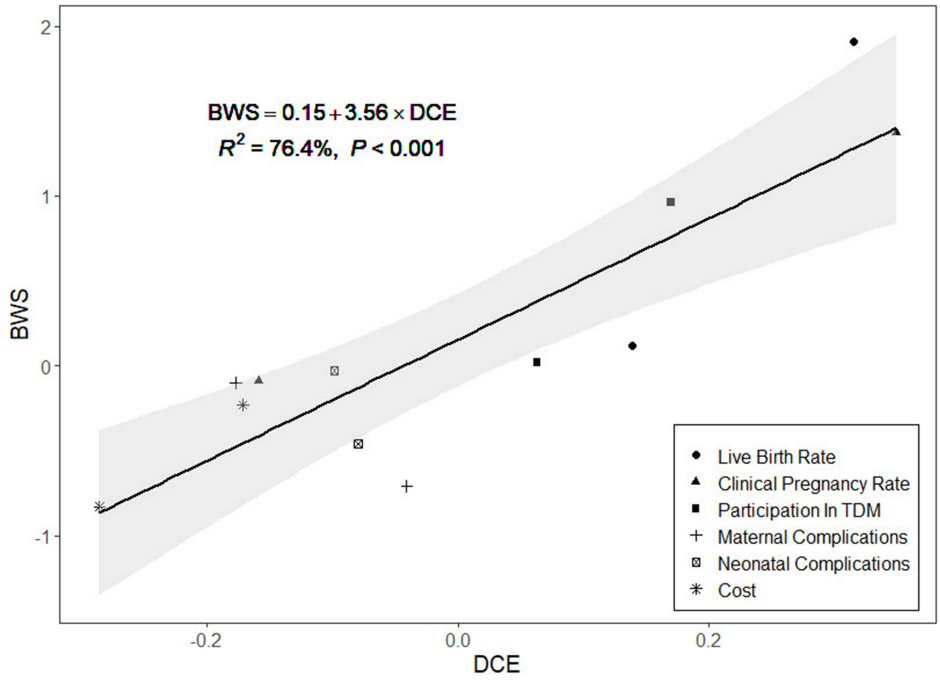

Figure 2 Plotting scatterplots of preference weight comparisons for BWS-2 and DCE.

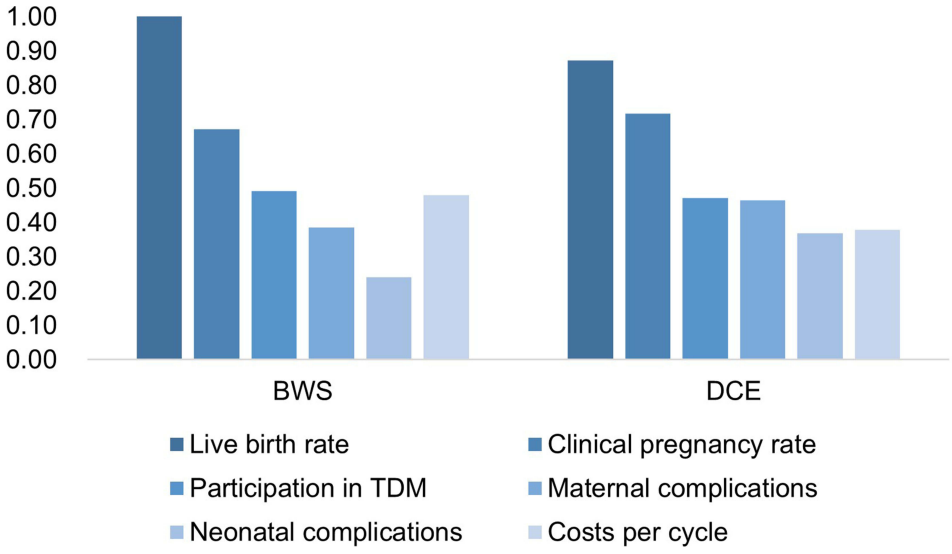

Figure 3 Relative importance of attributes for BWS-2 and DCE.

Comparison of Burden

In order to facilitate a comparative analysis of the two methods, this study required patients to complete both the DCE and BWS-2 questionnaires, as well as the assessment questions. To eliminate any potential order bias, the order in which the choice sets appeared was reversed in both versions. Table 5 presents the descriptive analyses, which indicate that patients were more likely to understand the DCE questions and preferred to complete the DCE questionnaire. A paired samples Wilcoxon test was conducted to compare respondents' scores on method difficulty. When the BWS-2 questions appeared first, patients demonstrated a higher mean difficulty in understanding the BWS-2 relative to the DCE. Additionally, a greater proportion of patients selected the DCE as their preferred measurement tool. Conversely, when

https://doi.org/10.2147/PPA.S501723

Table 5 Evaluation Questions for BWS-2 and DCE

	Version 1		Version 2		Overall	
	BWS 1st.	**DCE**	**BWS**	**DCE 1st.**	**BWS**	**DCE**
Easier to understand No. (%)	59(32.6)	122(67.4)	67(45.0)	82(55.0)	126(38.2)	204(61.8)
Prefer to fill out No. (%)	52(28.7)	129(71.3)	63(42.3)	86(57.7)	115(34.8)	215(65.2)
Difficulty score M±SD	4.71±2.52	3.72±2.65	4.52±2.61	4.50±2.83	4.63±2.56	4.07±2.76
Wilcoxon rank sum test	P<0.01		P=0.96		P<0.01	

the DCE appeared first, no significant difference in difficulty was observed. No significant difference was observed in the difficulty ratings of the BWS-2 by education level (p = 0.107). However, for the DCE choice set, there were statistically significant differences between the education groups (difficulty scores for respondents with different levels of education were 5.33,4.12,3.85 and 3.44 respectively; p=0.007). The same results are shown on the eTables 4 and 5. Respondents' choice of measurement tools was related to their educational qualifications.

Discussions

This study focuses on the differences in the estimation of preference weighting outcomes for infertility patients by BWS-2 and DCE in different assisted reproduction scenarios, along with patient problem assessment. It was found that the attribute levels of BWS-2 and DCE were similarly weighted, with patients preferring treatment options that were highly effective, had low side effects, were patient-centered, and were low-cost. There were only slight differences in the relative importance priority ranking of attributes in terms of individual attribute rankings. For patients, DCE questions were perceived to be less difficult to understand and were more likely to be filled out, whereas the difference in utility scale for DCE was smaller.

The strength of this study is that, although this is not the first study to focus on outcome differences between DCE and BWS-2, to our knowledge, no similar studies have been conducted within a specific infertility patient. Furthermore, this study is significant in light of the rising prevalence of infertility and the increasing trend towards younger age groups, a demographic that often exhibits positive attitudes towards triggering preferences and can act throughout the course of treatment. One of the primary findings of this study was the quantification of infertility patients' preference for medical treatment. The statistically significant preference results obtained from both the BWS-2 and DCE preference measurement tools, as well as the concordance, indicate that both preference measurement tools possess good validity in this study. This suggests that the DCE and BWS-2 research methodology are effective in revealing the preferences of infertility patients. The findings indicate that female infertility patients who are prepared to undergo or are undergoing ART derive greater utility from programs with higher success rates and lower risks. Furthermore, attributes related to effectiveness and safety, such as healthy live birth rate, pregnancy rate, and maternal complications, have the greatest impact on patient preference. A previous domestic study employing choice-based conjoint analysis to examine infertile patients' preferences for fertility care yielded comparable outcomes, indicating that patients placed the greatest value on their physician's attitude and treatment success.[32] Domestic patients are demonstrating a growing desire for enhanced quality of care, encompassing not only advances in health technology and health outcomes, but also the treatment process itself.[26] Patients anticipate more comprehensive doctor-patient interactions, more transparent treatment protocols, and so forth. In contrast to the Skedgel et al study,[24] the Chinese respondents all preferred treatments with a higher degree of shared decision making, but with a different emphasis on cost. In the past, the effect of preference on cost has not been significant. However, this study found that, similar to other countries, patients are starting to value cost attributes and prefer lower-cost treatment options. This may be due to the fact that China's fertility support policy is gradually improving and patients' out-of-pocket costs are becoming lower. Despite the fact that the technology employed in In Vitro Fertilization (IVF) in China has reached the world's advanced level, the invasive nature of this treatment means that it is not without certain side effects. In particular, the risk of neonatal complications was relatively less valued by older couples than by younger couples.[23] Differences in preferences may be due to the different respondents included in the

study. Both measurement tools in this study indicated that maternal complications were slightly more important than neonatal complications. However, in the last two studies,[33,34] the importance of preventing neonatal complications was higher when respondents included men or when couples were involved in the research at the same time. It is therefore imperative to gain an understanding of how patients weigh up the benefits and risks of treatment from their perspective if improvements are to be made to the patient treatment process and the quality of healthcare services enhanced. The emphasis should be placed on the patient experience and value-based healthcare, with the objective of enabling ART to transition from a consumer healthcare role to a core healthcare technology, with a particular focus on the patient experience in their services.

Despite adherence to international guidelines and recommendations, particularly those pertaining to DEC studies, in the construction of the experimental design. However, the specific choices and assumptions varied depending on the purpose of the study, which may affect the generalizability and results of this study.[35] These findings align with those of previous stated preference studies,[29,36,37] which demonstrated that different instruments exhibit comparable preference patterns. In this study, the importance of cost was found to vary in terms of its relative importance to other attributes. This may be attributed to the fact that relative importance is calculated as incremental differences rather than absolute numbers.[36] Secondly, this may be related to the context of health decision-making. Previously, the cost of ART was borne by the patient. However, since the introduction of the Active Fertility Policy by the Provincial Health Insurance Bureau in 2023, more than a dozen provinces and regions have included ART in their health insurance to implement reimbursement. It is possible that the implementation of the policy in Shanghai during the research process may have had an impact on the patients' preferences. Shanghai has included some therapeutic ART programs in the scope of payment under medical insurance and industrial injury insurance. For example, in the case of embryo culture programs, patients can receive up to three medical insurance reimbursements at the same medical institution. Finally, BWS-2 and DCE are compared, despite sharing the same random utility theory. However, the question format differs. BWS-2 is more inclined towards the sequential ordering of attribute levels, whereas DCE is in the form of combinations with trade-offs.[38] It can thus be supposed that patients may not arrive at the same decisions when making their choices. Additionally, studies have identified inconsistencies in the results obtained from the two tools. For instance, Whitty's review revealed a lack of concordance between DCE and BWS-2 findings.[35] The authors proposed that the BWS-2 may yield comparable results to the DCE only when assessing individual health preferences, rather than public preferences. Accordingly, this is closely related to the purpose of the study. The study design and the type of decision being made (eg, treatment choices, prioritization) may influence the decision-making process and the resulting choices, which may be affected by a variety of factors such as the study context and the disease population. This may result in different results. Furthermore, a significant benefit of BWS-2 over DCE is that by separating attribute weights and level preferences, researchers can obtain direct insights into preferences for attributes, eliminating the need for additional mathematical calculations. Additionally, policies that alter attributes may prove more effective.[39]

The second major finding of this study is the high degree of consistency between the BWS-2 and DCE preference results. However, contrary to expectations, an investigation into the willingness to complete the questionnaire and the difficulty of completing the questionnaire revealed that the sample population preferred to complete the DCE questions in this study. Previous studies have indicated that there may be some discrepancies in the acceptability of completing the two preference measurement tools, BWS-2 and DCE, among respondents in different sample populations.[6,40] In the present study, a positive correlation was observed between the level of education and the ratings of difficulty for the DCE questions. Patients with higher levels of education were more inclined to complete the DCE questionnaire. It is hypothesized that the difficulty in completing the questionnaire may be attributed to the cognitive burden resulting from the detachment of the choices from reality in more experimental designs and the overabundance of questions.[36,41] This may be partially related to the characteristics of interviewees. Furthermore, this study revealed that the sequence in which the choice sets were presented influenced respondents' comprehension. Despite the identical number of choice sets, the clear explanation of different choice tasks, along with patients' preconceived notions, may have contributed to the observed differences in comprehension. In conclusion, the results must be interpreted with caution. Ultimately, the change in utility associated with obtaining higher levels of the attributes differed between the two methods. In this study,

it was found that the utility scale of DCE is relatively limited and that patients' decision-making is not as stable Consequently, a larger sample size is required for DCE studies than for the other.[42]

There are some limitations to this study. Firstly, it was not possible to draw causal inferences due to the inherent limitations of cross-sectional studies. Secondly, the study employed a simple sampling method that only partially reflected the patients' views, and did not reveal the underlying reasons for their preferences. Further qualitative research, such as interviews, is necessary to gain a deeper understanding. Finally, it is imperative to exercise caution when interpreting the results of patient assessment questions, as they are subject to inherent limitations and potential biases.

Conclusions

In this study, the BWS-2 and DCE methodologies were employed to estimate attribute weights and level preferences for infertility patients' treatment decisions. The results demonstrated that patients' preferred decisions were stable and that the two measures yielded highly consistent results, with only slight differences in the calculated relative importance. However, patients indicated a preference for answering DCE questions and found the DCE approach to be straightforward and comprehensible. Further empirical research is required to gain a deeper understanding of the relative merits of these methods and to inform subsequent developments. The selection of an appropriate method, or the joint use of both DCE and BWS-2 methods, is essential for the purpose of informing health technology decisions, including the determination of utility values and risk-benefit ratios.

Data Sharing Statement

The data are available from the corresponding author on reasonable request.

Informed Consent Statement

Informed consent was obtained from all participants involved in the study.

Acknowledgment

We thank all the interviewers and respondents in our study.

Author Contributions

All authors have made significant contributions to the work reported, whether in conceptualization, study design, execution, data acquisition, analysis and interpretation, or all of these areas; have participated in drafting, revising or commenting on the article; have finalized the version to be published; have agreed on the journal to which the article will be submitted; and have agreed to take responsibility for all aspects of the work.

Funding

This work was supported by the National Natural Science Foundation of China (XL, grant number 72074123), (LZ, grant number 72304150) and Jiangsu Province Postgraduate Research and Practice Innovation Program (JC, grant number KYCX23_1901).

Disclosure

The authors declare that there are no competing interests in this work.

References

1. Viney R, Savage E, Louviere J. Empirical investigation of experimental design properties of discrete choice experiments in health care. *Health Econ*. 2005;14(4):349–362. doi:10.1002/hec.981
2. Mandeville KL, Lagarde M, Hanson K. The use of discrete choice experiments to inform health workforce policy: a systematic review. *BMC Health Serv Res*. 2014;14(1):367. doi:10.1186/1472-6963-14-367
3. McGrady ME, Pai ALH, Prosser LA. Using discrete choice experiments to develop and deliver patient-centered psychological interventions: a systematic review. *Health Psychol Rev*. 2021;15(2):314–332. doi:10.1080/17437199.2020.1715813

4. Benz HL, Saha A, Tarver ME. Integrating the voice of the patient into the medical device regulatory process using patient preference information. *Value Health*. 2020;23(3):294–297. doi:10.1016/j.jval.2019.12.005
5. Marley AAJ, Flynn TN. Probabilistic models of set-dependent and attribute-level best-worst choice. *J Math Psychol*. 2008;52(5):281–296.
6. Himmler S, Soekhai V, Van Exel J, Brouwer W. What works better for preference elicitation among older people? Cognitive burden of discrete choice experiment and case 2 best-worst scaling in an online setting. *J Choice Model*. 2021;38:100265. doi:10.1016/j.jocm.2020.100265
7. Whitty JA, Ratcliffe J, Chen G, Scuffham PA. Australian public preferences for the funding of new health technologies: a comparison of discrete choice and profile case best-worst scaling methods. *Med Decis Making*. 2014;34(5):638–654. doi:10.1177/0272989X14526640
8. Bridges JFP, Hauber AB, Marshall D, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR good research practices for conjoint analysis task force. *Value Health*. 2011;14(4):403–413. doi:10.1016/j.jval.2010.11.013
9. Hauber AB, González JM, Groothuis-Oudshoorn CGM, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health*. 2016;19(4):300–315. doi:10.1016/j.jval.2016.04.004
10. Reed Johnson F, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value Health*. 2013;16(1):3–13. doi:10.1016/j.jval.2012.08.2223
11. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best–worst scaling: what it can do for health care research and how to do it. *J Health Econ*. 2007;26 (1):171–189. doi:10.1016/j.jhealeco.2006.04.002
12. Cheung KL, Wijnen BFM, Hollin IL, et al. Using best–worst scaling to investigate preferences in health care. *PharmacoEconomics*. 2016;34 (12):1195–1209. doi:10.1007/s40273-016-0429-5
13. Qiao J, Wang Y, Li X, et al. A Lancet Commission on 70 years of women's reproductive, maternal, newborn, child, and adolescent health in China. *Lancet*. 2021;397(10293):2497–2536. doi:10.1016/S0140-6736(20)32708-2
14. Tatum M. China's fertility treatment boom. *Lancet*. 2020;396(10263):1622–1623. doi:10.1016/S0140-6736(20)32475-2
15. Zhou Z, Zheng D, Wu H, et al. Epidemiology of infertility in China: a population-based study. *BJOG*. 2018;125(4):432–441. doi:10.1111/1471-0528.14966
16. Bornstein M, Gipson JD, Failing G, Banda V, Norris A. Individual and community-level impact of infertility-related stigma in Malawi. *Soc Sci Med*. 2020;251:112910. doi:10.1016/j.socscimed.2020.112910
17. Kong F, Wang Y, Li R, Qiao J. Assisted reproductive technology in China: introduction to the special issue. *Hum Reprod*. 2023;38(Supplement_2): ii1–ii2. doi:10.1093/humrep/dead215
18. Adamson GD, Zegers-Hochschild F, Dyer S. Global fertility care with assisted reproductive technology. *Fertil Steril*. 2023;120(3 Pt 1):473–482. doi:10.1016/j.fertnstert.2023.01.013
19. Keller E, Botha W, Chambers GM. What features of fertility treatment do patients value? Price elasticity and willingness-to-pay values from a discrete choice experiment. *Appl Health Econ Health Policy*. 2023;21(1):91–107. doi:10.1007/s40258-022-00764-7
20. Van Der Horst DEM, Garvelink MM, Bos WJW, Stiggelbout AM, Pieterse AH. For which decisions is shared decision making considered appropriate? – a systematic review. *Patient Educ Counsel*. 2023;106:3–16. doi:10.1016/j.pec.2022.09.015
21. Harrison C, Gameiro S, Boivin J. Patient willingness, preferences and decision-making about planning for three complete cycles of IVF/ICSI treatment. *Hum Reprod*. 2021;36(5):1339–1352. doi:10.1093/humrep/deab040
22. Nachtigall RD, Dougall KM, Lee M, Harrington J, Becker G. What do patients want? Expectations and perceptions of IVF clinic information and support regarding frozen embryo disposition. *Fertil Steril*. 2010;94(6):2069–2072. doi:10.1016/j.fertnstert.2010.02.023
23. Weiss NS, Schreurs AMF, Van Der Veen F, et al. Women's perspectives on ovulation induction with or without IUI as treatment for normogonadotrophic anovulation: a discrete choice experiment. *Hum Reprod Open*. 2017;2017(3). doi:10.1093/hropen/hox021
24. Skedgel C, Ralphs E, Finn E, Markert M, Samuelsen C, Whitty JA. How do people with experience of infertility value different aspects of assistive reproductive therapy? Results from a multi-country discrete choice experiment. *Patient*. 2022;15(4):459–472. doi:10.1007/s40271-021-00563-7
25. Klitzman R. Impediments to communication and relationships between infertility care providers and patients. *BMC Women's Health*. 2018;18(1):84. doi:10.1186/s12905-018-0572-6
26. Geng L, Shi Z, Chai X, et al. Patient and clinician perspectives on shared decision-making in infertility treatment: a qualitative study. *Patient Educ Counsel*. 2023;116:107948. doi:10.1016/j.pec.2023.107948
27. De Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *Patient*. 2015;8(5):373–384. doi:10.1007/s40271-015-0118-z
28. Botha W, Donnolley N, Shanahan M, Norman RJ, Chambers GM. Societal preferences for fertility treatment in Australia: a stated preference discrete choice experiment. *J Med Econ*. 2019;22(1):95–107. doi:10.1080/13696998.2018.1549055
29. Potoglou D, Burge P, Flynn T, et al. Best–worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Soc Sci Med*. 2011;72(10):1717–1727. doi:10.1016/j.socscimed.2011.03.027
30. Veldwijk J, Lambooij MS, de Bekker-Grob EW, Smit HA, de Wit GA. The effect of including an opt-out option in discrete choice experiments. *PLoS One*. 2014;9(11):e111805. doi:10.1371/journal.pone.0111805
31. Wang Y, Zhai P, Jiang S, Li C, Li S. Blood donors' preferences toward incentives for donation in China. *JAMA Netw Open*. 2023;6(6):e2318320. doi:10.1001/jamanetworkopen.2023.18320
32. Cai QF, Wan F, Dong XY, et al. Fertility clinicians and infertile patients in China have different preferences in fertility care. *Hum Reprod*. 2014;29 (4):712–719. doi:10.1093/humrep/deu023
33. Abdulrahim B, Scotland G, Bhattacharya S, Maheshwari A. Assessing couples' preferences for fresh or frozen embryo transfer: a discrete choice experiment. *Hum Reprod*. 2021;36(11):2891–2903. doi:10.1093/humrep/deab207
34. Hartman H, Kermanshahi N, Matzkin E, Keyser EA, Gianakos AL. Decreased fertility awareness amongst surgeons and surgical trainees and potential role of formal fertility education. *J Surg Educ*. 2024;81(7):947–959. doi:10.1016/j.jsurg.2024.03.011
35. 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. *Patient*. 2018;11(3):301–317. doi:10.1007/s40271-017-0288-y
36. Van Dijk JD, Groothuis-Oudshoorn CGM, Marshall DA, IJzerman MJ. An empirical comparison of discrete choice experiment and best-worst scaling to estimate stakeholders' risk tolerance for hip replacement surgery. *Value Health*. 2016;19(4):316–322. doi:10.1016/j.jval.2015.12.020
37. Severin F, Schmidtke J, Mühlbacher A, Rogowski WH. Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments. *Eur J Hum Genet*. 2013;21(11):1202–1208. doi:10.1038/ejhg.2013.36

38. Krucien N, Watson V, Ryan M. Is best–worst scaling suitable for health state valuation? A comparison with discrete choice experiments. *Health Econ*. 2017;26(12):e1–e16. doi:10.1002/hec.3459
39. Flynn TN, Louviere JJ, Peters TJ, Coast J. Estimating preferences for a dermatology consultation using best-worst scaling: comparison of various methods of analysis. *BMC Med Res Methodol*. 2008;8(1):76. doi:10.1186/1471-2288-8-76
40. Huynh E, Coast J, Rose J, Kinghorn P, Flynn T. Values for the ICECAP-supportive care measure (ICECAP-SCM) for use in economic evaluation at end of life. *Soc Sci Med*. 2017;189:114–128. doi:10.1016/j.socscimed.2017.07.012
41. Rogers HJ, Marshman Z, Rodd H, Rowen D. Discrete choice experiments or best-worst scaling? A qualitative study to determine the suitability of preference elicitation tasks in research with children and young people. *J Patient Rep Outcomes*. 2021;5(1):26. doi:10.1186/s41687-021-00302-4
42. Soekhai V, Donkers B, Johansson JV, et al. Comparing outcomes of a discrete choice experiment and case 2 best-worst scaling: an application to neuromuscular disease treatment. *Patient*. 2023;16(3):239–253. doi:10.1007/s40271-023-00615-0