



How Do People with Experience of Infertility Value Different Aspects of Assistive Reproductive Therapy? Results from a Multi-Country Discrete Choice Experiment

Chris Skedgel¹ · Eleanor Ralphs² · Elaine Finn² · Marie Markert⁴ · Carl Samuelsen⁴ · Jennifer A. Whitty³

Accepted: 9 November 2021
© The Author(s) 2021

Abstract

Objectives Assistive reproductive therapies can help those who have difficulty conceiving but different forms of assistive reproductive therapies are associated with different treatment characteristics. We undertook a large, multinational discrete choice experiment to understand patient preferences for assistive reproductive therapies.

Methods We administered an online discrete choice experiment with persons who had experience with subfertility or assistive reproductive therapies in the USA, UK, the Nordic region (Denmark, Norway, Sweden, Finland), Spain, and China. Attributes encouraged trade-offs between effectiveness, risk of adverse effects, treatment (dis)comfort, (in)convenience, cost per cycle and shared decision making. We used multinomial logit and mixed-logit models to estimate preferences and attribute importance by country/region, and estimated willingness to pay for changes in attribute levels.

Results A total of 7565 respondents participated. Mixed logit had a better fit than multinomial logit across all samples. Preferences moved in expected directions across all samples, but the relative importance of attributes differed between countries. Willingness to pay was greatest for improvements in effectiveness and a greater degree of shared decision making, and we observe a substantial ‘option value’ independent of treatment characteristics. Unexpectedly, preferences over cost were insignificant in the Chinese sample, limiting the use of willingness to pay in this sample.

Conclusions Respondents balanced concerns for effectiveness with other considerations, including the cost and (dis)comfort of treatment, and the degree of shared decision making, but there is also substantial option value independent of treatment characteristics, demonstrating value of assistive reproductive therapies to individuals with experience of subfertility. We hypothesise that price insensitivity in the Chinese sample may reflect a degree of social desirability bias.

1 Introduction

Cultural, demographic and other trends over recent decades have led to later childbearing as well as increasing obesity rates, a rise of sexually transmitted diseases and decreasing sperm quality [1–3]. Together, these have meant that an increasing number of prospective parents are experiencing subfertility, defined as an inability to achieve a clinical pregnancy after trying for more than 12 months [4]. A recent

review found that in more developed countries, 12-month infertility rates ranged from 3.5 to 16.7%, and that 40–70% of these cases sought medical treatment for infertility [5].

Assistive reproductive therapies (ART) can help individuals or couples who have difficulty conceiving naturally to get pregnant and carry a baby to term. However, while ART improve a couple’s chance of conceiving a child, different forms of treatment are associated with differing effectiveness, risks and convenience. These characteristics can affect patient preferences for specific forms of ART. An understanding of how patients prioritise and trade-off between the positive and negative aspects of treatments is essential to ensuring that their care is aligned with their preferences and therefore provides the greatest value. Research has also suggested that physicians and patients often differ in their view of the most important characteristics of treatment, including infertility treatments [6–8]. A misaligned understanding of patient preferences can undermine shared decision making, as providers may emphasise aspects of treatment that are less

✉ Chris Skedgel
cskedgel@ohe.org

¹ Office of Health Economics, Southside, 7th Floor, 105 Victoria Street, London SW1E 6QT, UK

² IQVIA, London, UK

³ Health Economics Group, University of East Anglia, Norwich, UK

⁴ Ferring Pharmaceuticals, Copenhagen, Denmark

Key Points for Decision Makers

This study provides evidence from large multi-national samples to generalise the results of previous smaller scale research around patient preferences for assistive reproductive therapies.

Effectiveness of treatment was the most (or second most) important attribute across all samples but we do not see a preference for effectiveness to the exclusion of other considerations, or a wide gap between effectiveness and other attributes in terms of relative importance.

Respondents placed significant value on access to treatment, reflected in the 'option value' of treatment, but also had a substantial willingness to pay for improvements in the effectiveness of treatment, a greater degree of shared decision making, and among some respondents, less discomfort in treatment.

The Chinese sample was insensitive to the cost of treatment in their choices, although their preferences across the other attributes were broadly similar to the other samples. We hypothesise that this result may reflect some social desirability bias that discouraged respondents in this sample from considering cost in their choice of treatment.

important to patients, or misunderstand acceptable trade-offs between different aspects of treatment.

Previous research around preferences for fertility treatment has found that effectiveness, in terms of the probability of a live birth, is typically (but not invariably) the single most important characteristic of a treatment to patients, but that they are willing to accept some reduction in the probability of a successful live birth to improve other aspects of treatment. These other aspects include lower risks of adverse events, more convenient modes of administration and a greater degree of shared decision making [6, 7, 9–11]. Most of these studies, though, have been relatively small by the current standards of stated preference research [12], usually 200 respondents or less, and are reflective of a single country. This limits the ability to generalise the results across patients or to understand how treatment preferences might change in different national contexts. Additionally, many studies in this area have not included cost as an attribute, which is arguably a significant shortcoming given that many patients must finance their own fertility treatment in many countries [13–15].

Our primary objective here is to quantify the relative importance of different aspects of fertility treatment in a

more generalisable context, including in terms of willingness to pay (WTP) for different characteristics of ART. As such, we undertook a large discrete choice experiment (DCE) with persons who had experience with subfertility or ART in five countries/regions: USA, UK, the Nordic region (Denmark, Norway, Sweden and Finland), Spain and China.

2 Methods

We administered an online survey in two sections. In the first section, respondents were asked about their demographic characteristics, experience with subfertility, their attitudes towards infertility and their willingness to contribute to a publicly funded ART programme. These results are reported elsewhere [16]. In the second section, respondents who indicated that they had tried for more than 12 months to have a baby or had sought medical treatment to get pregnant were presented a series of DCE tasks to elicit their preferences over different aspects of ART. The survey itself is available as Electronic Supplementary Material (ESM).

A DCE is a quantitative approach to eliciting individuals' preferences over health states. Respondents are asked to choose their most preferred option from a choice set of two or more alternatives described in terms of a common set of attributes and differing attribute levels. This methodology has previously been applied in the context of infertility and ART [6, 7, 9–11] as well as in other healthcare contexts [12, 17].

2.1 Design of the DCE

The attributes included in the DCE were derived through reference to previous DCEs and other preference studies in this area [6, 7, 9, 11, 18, 19]. Attributes were selected to encourage consideration of trade-offs between treatment effectiveness, risk of adverse effects, treatment (dis)comfort and (in)convenience, and cost per ART cycle. We also sought to understand the importance of patient centredness or shared decision making in treatment decisions relative to other attributes.

Attribute levels were selected by the authors to cover the range of salient levels, with reference to the previous DCEs mentioned above along with unstructured input and review from a clinical expert in reproductive medicine. The midpoint cost per cycle in each country was based on indicative costs from IVF-Worldwide [20], updated to 2020 costs using country-specific price indices. IVF-Worldwide cost estimates were calculated according to a methodology reported by Collins [21] and includes country-specific costs associated with initial consultation, basic in vitro fertilisation treatment, intracytoplasmic sperm injection, hormonal drugs, embryo freezing, other investigations and regulatory

fees. Estimates do not include productivity or other indirect costs. Upper and lower levels were defined as $\pm 40\%$ of the midpoint, reflecting the reported variation in UK cost per cycle. [20] Costs were presented to respondents in country-specific currencies but converted to Euro using XE.com historical exchange rates at the time of data analysis for comparability between study countries. The attributes and attribute levels in the experimental design are shown in Table 1, and the country-specific costs are shown in Table 2.

We used Ngene™ software (ChoiceMetrics Pty Ltd; Sydney, New South Wales, Australia) version 1.2.1 to generate a D-efficient fractional factorial experimental design, based on a main-effects model focusing on the independent effect of each attribute on choice. We assumed non-informative priors in developing the design. We produced a 36-set design with two treatment alternatives in each set and included a fixed ‘no treatment’ alternative in each choice task, with all attribute levels set to zero (i.e. no change from the current state). An example DCE task is shown in Box 1. The attributes and levels presented in the tasks were described to respondents in the introduction to the DCE, included as part of the survey in the ESM.

2.2 Survey

Samples were recruited from general population survey panels maintained by Dynata™ in USA, UK, Denmark, Norway, Sweden, Finland, Spain and China. These countries were chosen to represent a diverse cross-section of cultural attitudes and preferences towards infertility and ART. Because

Table 1 Attributes and levels in the discrete choice experiment

Attribute	Levels
Effectiveness (probability of success)	10%; 25%; 40%
Risk of complications	2%; 5%; 8%
Discomfort (hypo-estrogenemia)	None or mild; strong
Shared decision making	None; some; full
Daily injections	1; 3; 5
Cost per cycle ^a	Low; medium; high

^aCost per cycle varied by country. Country-specific costs are shown in Table 2

Table 2 Indicative cost per cycle by study country

Cost	UK	USA	Spain	Denmark	Finland	Norway	Sweden	China
Low (−40%)	£2400	\$8700	€3300	kr 29,100	€1800	kr 20,700	kr 50,100	¥12,600
Indicative (medium)	£4000	\$14,500	€5500	kr 48,500	€3000	kr 35,000	kr 83,500	¥21,000
High (+40%)	£5600	\$20,300	€7700	kr 67,900	€4200	kr 48,300	kr 116,900	¥29,400

Indicative costs include initial consultation, basic in-vitro fertilization, intracytoplasmic sperm injection, hormonal drugs, embryo freezing, other investigations and regulatory fees [20]

of the small populations of the individual Nordic countries, participants from these countries were pooled into a combined Nordic sample. Nationally representative samples in terms of age and sex were recruited in each country/region and supplemented by an ‘over-sample’ of reproductive age respondents to ensure sufficient statistical power for the DCE phase of the study. This ‘over-sample’ is not nationally representative as it is based on individuals with self-reported experience of subfertility or ART. The supplementary sample size was informed by recent practice in DCE elicitation [12].

Individuals who had previously registered with Dynata™ received an e-mail inviting them to learn more about this study. An accompanying link took them to an online participant information sheet (PIS), which outlined the purpose of the study and provided a link to the questionnaire. The PIS, questionnaire, and the statistical analysis plan were reviewed and approved by the University of East Anglia Faculty of Medicine and Health Science Ethics Committee, Norwich UK (reference 201819-090).

Each respondent saw 11 choice sets: ten unique sets plus one repeated set to test respondent consistency. The unique sets were selected ‘dynamically’: the ten sets with the fewest number of responses to that point in the data collection were selected from the full experimental design to ensure that each set was seen a similar number of times across all respondents. In the repeated set, an earlier task was re-presented with the order of two treatment alternatives reversed. In all cases, the third task was reversed and re-presented as the eighth task. Respondents who did not choose the same alternative (including ‘no treatment’) in both tasks were flagged as potentially non-attentive [22]. We recorded completion times and flagged respondents who completed the questionnaire in less than half the median completion time for their country. All respondents were included in the primary analysis but respondents flagged as both fast and inconsistent were excluded in a sensitivity analysis.

Respondents were asked for demographic details including their age group, highest level of education and income category. Each version of the questionnaire presented five income categories. In the UK, these categories were presented in £15,000 intervals (< 15,000; 15,000–30,000; 30,000–45,000; 45,000–60,000; > 60,000), and the other versions used roughly equivalent intervals in local currency.

Box 1 Sample discrete choice experiment task

Treatment 1	Treatment 2	No treatment
Treatment will result in 25 pregnancies in 100 couples	Treatment will result in 10 pregnancies in 100 couples	No improved chance of pregnancy
5 women in 100 will have moderate or severe complications	2 women in 100 will have moderate or severe complications	No risk of treatment-related complications
Side effects of treatment are mild	Side effects of treatment are strong	No treatment-related side effects
You will have full involvement in decisions about your treatment	You will have no involvement in decisions about your treatment	No physician contact
You will require 5 injections per day	You will require 3 injections per day	No injections per day
Treatment will cost ££££ ^a per cycle	Treatment will cost ££££ ^a per cycle	No treatment-related cost
o I prefer Treatment 1	o I prefer Treatment 2	o I prefer no treatment

^a££££ was replaced with a country-specific currency symbol and cost per cycle

A small convenience sample ($N = 25$) was recruited from each country/region to pilot the full survey. Respondents were asked to rate the difficulty and length of the survey on a 5-point Likert scale, from very easy/short to much too difficult/long. The pilot identified an issue around the currency symbols presented to respondents in China; this was corrected in the final version. Likert ratings of the length and difficulty of the survey did not flag concerns: 7% of pilot respondents found the survey ‘long’ or ‘very long’ and 9% found it ‘difficult’ or ‘very difficult’.

2.3 Statistical Analysis

Prior to modelling DCE responses, we generated descriptive statistics of the frequency of ‘no treatment’ choices and tested for non-trading or dominant preferences to confirm the theoretical validity of the elicitation. Respondents with a dominant preference *always* choose the alternative that maximises or minimises the level of a single attribute, such as treatment effectiveness, without regard to the level of other attributes such as cost or discomfort. Strictly dominant preferences are inconsistent with the theory of compensatory decision making that underlies DCE methods, and an ‘excessive’ proportion of dominant preferences may invalidate a DCE. However, such preferences are not ‘irrational’, and they are almost impossible to definitively identify in a fractional factorial design where respondents see only a subset of all possible attribute-level combinations [23, 24]. As such, we note the proportion of respondents with potentially dominant preferences but do not exclude these respondents from the analysis.

Discrete choice experiment responses were analysed by country/region. Cost was included in the analysis as a continuous variable and all other attributes were effects coded to allow for non-linear preferences over the levels of the different attributes. The middle level of each attribute was used as the reference level, except for the

two-level discomfort attribute, where ‘mild’ was used as the reference.

We specified an additive, main effects utility function for the treatment alternatives A and B, and specified the ‘no treatment’ alternative C as the reference alternative:

$$U(A) = \alpha_A + \beta_1 \times \text{Effect10} + \beta_2 \times \text{Effect40} + \beta_3 \times \text{Complications2} + \beta_4 \times \text{Complications8} + \beta_5 \times \text{Discomfort.Strong} + \beta_6 \times \text{SharedDM.None} + \beta_7 \times \text{SharedDM.Full} + \beta_8 \times \text{Injections1} + \beta_9 \times \text{Injections5} + \beta_{10} \times \text{CycleCost.Euros}/100,$$

$$U(B) = \alpha_B + \beta_1 \times \text{Effect10} + \beta_2 \times \text{Effect40} + \beta_3 \times \text{Complications2} + \beta_4 \times \text{Complications8} + \beta_5 \times \text{Discomfort.Strong} + \beta_6 \times \text{SharedDM.None} + \beta_7 \times \text{SharedDM.Full} + \beta_8 \times \text{Injections1} + \beta_9 \times \text{Injections5} + \beta_{10} \times \text{CycleCost.Euros}/100,$$

$$U(C) = 0,$$

where α_A and α_B are treatment-specific constants, representing the utility of treatment relative to no treatment, independent of attribute levels. We averaged these treatment-specific constants to represent the value of having treatment options, independent of the characteristics of those treatments. We refer to this value as ‘option value’.

In the first instance, we used separate multinomial logit models to estimate the part-worth utilities of each attribute level for each country/region. To allow for unobserved heterogeneity, we also tested random parameters, or mixed-logit models [25]. We assigned a normal distribution to all parameters except cost and generated 1000 Halton draws. We modelled cost as a deterministic parameter to facilitate estimates WTP for changes in attribute levels.

Where heterogeneity in the random coefficients was statistically significant, we used the Krinsky–Robb approach to estimate non-parametric 95% confidence intervals around the point estimate, based on the 2.5th and 97.5th percentiles

of the random coefficient draws [26]. Where heterogeneity was not statistically significant, we used the standard error of the point estimate. The Akaike Information Criterion was used to compare the fit of the multinomial logit and mixed-logit models.

To understand heterogeneity in preferences by respondent characteristics, we estimated a series of models including the main effects as above as well as an interaction term between each of the main effects and specific characteristic flags (e.g. $\beta_1 \times \text{Effect10} + \beta_2 \times \text{Effect10} \times \text{Female}$). Each subgroup was estimated separately. These interaction terms capture the difference in part-worth utilities between the specified subgroup and the remainder of the sample. We combined the national samples into a single dataset to test the impact of respondent characteristics other than nationality, specifically: ‘fast’ vs ‘non-fast’ responders; female vs male individuals; higher (quintiles 4 and 5) vs lower income; inconsistent vs consistent in the repeated task; not in a long-term relationship vs in a long-term relationship; and received ART vs no ART. Note that we rescaled the cost attribute to better highlight differences between subgroups. We also tested the impact of excluding respondents flagged as jointly ‘fast and inconsistent’ from the national samples in a sensitivity analysis.

The relative importance of the main effects for each sample was estimated as the absolute difference in the part-worth utility of the most preferred and least preferred levels of each attribute, as a share of the sum of differences across all attributes. Under this approach, attributes with a greater absolute difference in utility are relatively more important than attributes with a smaller absolute difference in utility. [27]

Finally, the implied WTP for a change in attribute levels was estimated using the cost attribute to reframe the part-worth utilities in terms of Euro. We estimated WTP using Small and Rosen’s compensating variation approach [28]:

$$\text{WTP}_x = \frac{1}{-\beta_{\text{Cost}}} (v_{x1} - v_{x0}),$$

where β_{Cost} is the coefficient on the cost parameter and v_{x0} and v_{x1} are part-worth utilities before and after a change in the level of attribute x . Given our main effects specification ($v_{x1} - v_{x0}$) is equivalent to $(\beta_{x1} - \beta_{x0})$, where β_{x0} is the reference level of attribute x and β_{x1} is the new level. We estimated WTP for the ‘option value’ of treatment as $\frac{\bar{\alpha}}{-\beta_{\text{Cost}}}$.

All analyses were conducted in R statistical software, version 4.0.5. [29] The MLOGIT [30] package was used to model choices, and the GGLOT2 [31] and GGPUBR [32] packages were used to produce the figures.

3 Results

The survey was administered in February 2020 and age–sex quotas for all samples were filled within 2 weeks of sending the first invitations. The characteristics of 7565 respondents who indicated that they had tried to have a baby and experienced 12 months or more of infertility, received medical assistance to try to get pregnant, or both, are shown in Table 3 by country/region. For most countries/regions, the number of participants peaked between 31 and 45 years of age, although the Chinese sample was slightly younger than the others, peaking between 26 and 35 years of age. The largest proportion of the Chinese and Nordic samples, 52% and 35%, respectively, were in the third of the five income categories, corresponding with a UK income of £30,000–45,000. The largest proportions of the Spanish (51%) and UK (33%) samples were in the second quintile (corresponding with £15,000–30,000), and the largest proportion of the USA sample (24%) was in the fourth quintile (corresponding with £45,000–60,000).

Response behaviours are summarised in Table 4. Median DCE completion times ranged from 1½ to 2½ minutes, and approximately 20% of respondents had a completion time of less than half their country-specific median. Inconsistency in the repeated task was between 32 and 40% across the countries/regions in the survey. The Nordic (13%) and UK (13%) samples had significantly higher proportions of joint ‘fast and inconsistent’ respondents than China (8%), USA (10%) and Spain (9%).

Twelve per cent of all choices were for ‘no treatment’, and analysis of variance showed that the proportion of ‘no treatment’ choices was significantly lower in China (7.1%) relative to other regions (13.6–18.4%). Six hundred and six respondents (8.0%) chose ‘no treatment’ in a majority of the tasks they saw (six or more ‘no treatment’ choices out of 11 tasks) and 256 (3.4%) chose ‘no treatment’ in all tasks. Logistic regression showed a statistically significant association between choosing no treatment in a majority of tasks and increasing respondent age, and that female individuals were significantly more likely to choose no treatment than male individuals, as were respondents in the lowest income category relative to the middle-income category.

There was some evidence of dominant preferences around effectiveness and discomfort. Of all respondents, 7.8% and 5.7% always chose the alternative that maximised the level of effectiveness or minimised the level of discomfort, respectively. The proportion of respondents with a dominant preference for these two attributes was substantially and significantly greater than for the other attributes (< 1% for all other attributes). Analysis of variance showed that the proportion of respondents with a potentially dominant preference for maximising effectiveness was significantly greater in the UK

Table 3 Respondent counts and characteristics by country/region

Country/region	Total respondents	Female ^b (%)	Married or long-term relationship ^b (%)	Tried ≥ 12 months ^b (%)	Received medical assistance ^b (%)
China	2571	57	99	97	67
Nordic countries ^a	829	58	84	93	52
Spain	1688	52	92	89	63
UK	1260	59	90	96	48
USA	1217	59	88	90	54

All values are conditional on having tried to have a baby

^aDenmark 201; Finland 173; Norway 158; Sweden 297

^bProportions excluded respondents who declined to answer

Table 4 Response behaviours by country/region

Country/region	Median DCE completion time (minutes)	'Fast completers' (less than half median time) (%)	Inconsistent in repeated task (%)	Fast and inconsistent responder (%)
China	1:28	17	40	9
Nordic countries	2:27	23	35	9
Spain	1:56	19	32	10
UK	1:40	19	36	5
USA	1:31	17	38	6

DCE discrete choice experiment

(11.6%) than other countries, with no significant differences between other regions (5.9–8.1%). Conversely, the proportion with a potentially dominant preference for minimising discomfort was significantly greater in China (10.4%) and Spain (8.0%) relative to other countries (1.3–3.4%).

3.1 Preference Modelling

The mixed-logit models had the best fit for each country/region. Country-specific coefficients, or “part-worth utilities”, are presented in Table 5 and illustrated in Fig. 1. In this figure, an upward sloping line indicates that a higher level of the attribute was preferred, whilst downward sloping preference indicates that the lower level of the attribute was preferred. Most point estimates were statistically significant with the exception of preferences over the number of daily injections in some samples and, most notably, the cost attribute in the Chinese sample. We observed significant heterogeneity in preferences over effectiveness, discomfort and shared decision making. The confidence interval around discomfort crossed zero in all samples, but the other attributes remained significant for most samples.

Heterogeneity between subgroups is summarised in Fig. 2, with point estimates and 95% confidence intervals as well as the proportion of respondents in each subgroup. The greatest deviations were between fast (less than half

median completion time) and ‘non-fast’ respondents, and between consistent and inconsistent respondents. We return to the issue of these respondents in a sensitivity analysis below. Among female respondents, preferences for different attribute levels moved in the same direction as male respondents but tended to be relatively stronger. For example, female respondents derived greater positive utility from the higher level of effectiveness and greater negative utility from a lower level of effectiveness relative to the remainder of the sample. There are examples of statistically significant divergences in the strength of preference amongst the other subgroups, but these are relatively small in absolute terms.

Table 5 and Fig. 3 show the relative contribution of each attribute to overall utility, conditional on the ranges presented to respondents. Effectiveness was the most important attribute in most countries and the number of injections was the least important attribute in most countries. The importance of cost was highly variable across samples, from statistically insignificant in China to almost as important as effectiveness in USA. The importance of (dis)comfort was also variable, from the most important factor in China to relatively unimportant in UK and USA. The degree of shared decision making was more important than cost in the Nordic sample and was also relatively important in UK and USA, but relatively unimportant in China.

Table 5 Model coefficients and p-values by sample

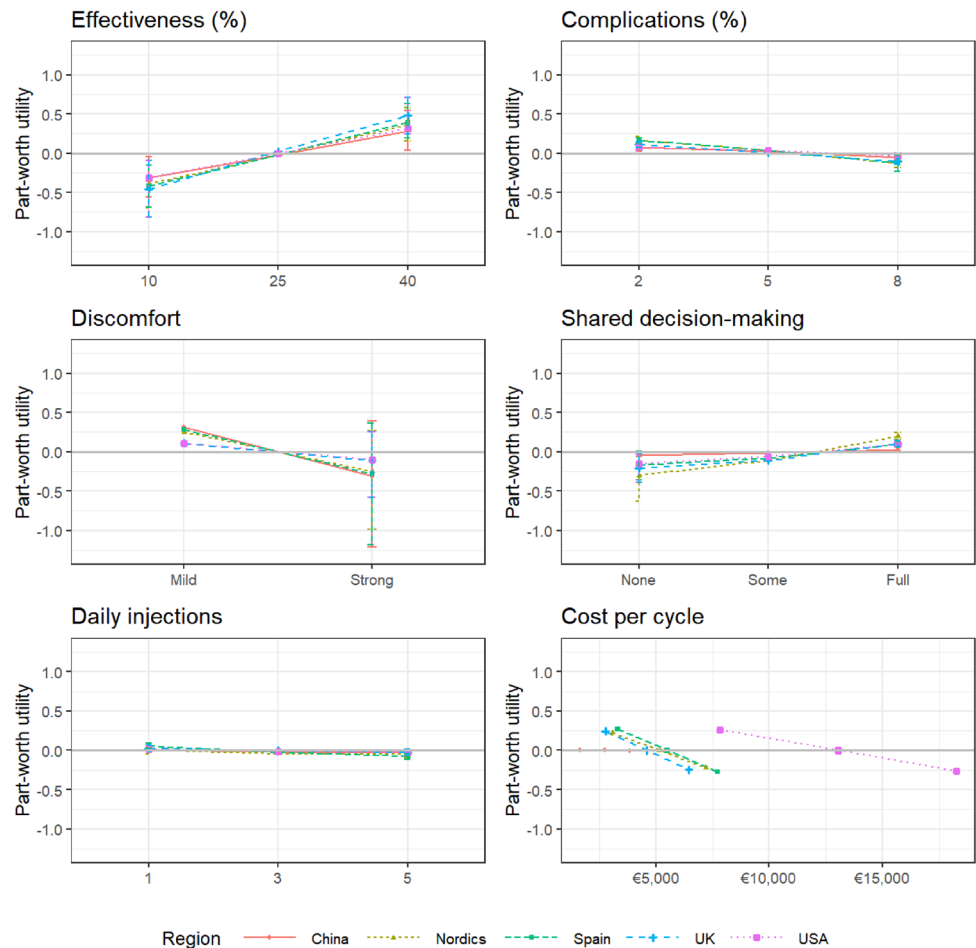
Parameter	Distribution	China	Nordic countries	Spain	UK	USA
Treatment A constant	Deterministic	1.5984	1.0902	1.5408	1.4943	1.2924
[p value]		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Treatment B constant	Deterministic	1.8755	1.3085	1.7742	1.7299	1.5350
[p value]		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
10% effectiveness flag	Normal	-0.3073	-0.3864	-0.4167	-0.4566	-0.3152
[p value]		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
40% effectiveness flag	Normal	0.2836	0.3610	0.3943	0.4778	0.3087
[p value]		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
2% complication risk flag	Normal	0.0730	0.1665	0.1655	0.1107	0.0656
[p value]		<0.0001	<0.0001	<0.0001	<0.0001	0.0003
8% complication risk flag	Normal	-0.0484	-0.1323	-0.1255	-0.1054	-0.0332
[p value]		0.0002	<0.0001	<0.0001	<0.0001	0.0649
Strong discomfort flag	Normal	-0.3126	-0.2499	-0.2862	-0.1085	-0.1071
[p value]		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
No shared DM flag	Normal	-0.0406	-0.3048	-0.1732	-0.2139	-0.1524
[p value]		0.0017	<0.0001	<0.0001	<0.0001	<0.0001
Full shared DM flag	Normal	0.0272	0.1956	0.0918	0.1047	0.0901
[p value]		0.0369	<0.0001	<0.0001	<0.0001	<0.0001
1 daily injection flag	Normal	0.0083	0.0025	0.0571	0.0236	0.0152
[p value]		0.5240	0.9160	0.0006	0.1966	0.3933
5 daily injections flag	Normal	-0.0271	-0.0465	-0.0775	-0.0255	-0.0258
[p value]		0.0374	0.0467	<0.0001	0.1649	0.1487
Cost per cycle (Euro)/100	Fixed	-0.0001	-0.0108	-0.0122	-0.0134	-0.0050
[p value]		0.9339	<0.0001	<0.0001	<0.0001	<0.0001
sd.effect_10%		0.2908	0.3221	0.3003	0.3620	0.2199
[p value]		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
sd.effect_40%		0.2834	0.2588	0.2674	0.2747	0.2253
[p value]		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
sd.complications_2		0.1169	-0.0041	0.0892	0.0216	0.0010
[p value]		0.0364	0.9971	0.3362	0.9546	0.9994
sd.complications_8		-0.0170	0.0598	0.1766	0.0800	0.0015
[p value]		0.9546	0.7585	0.0005	0.4827	0.9992
sd.discomfort_Strong		0.5560	0.4161	0.5059	0.3214	0.2510
[p value]		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
sd.sharedDM_None		0.0969	0.0832	0.0036	-0.0020	0.0061
[p value]		0.1427	0.5467	0.9949	0.9980	0.9941
sd.sharedDM_Full		-0.0510	0.3298	-0.2383	0.2308	0.1895
[p value]		0.6737	<0.0001	<0.0001	<0.0001	0.0001
sd.injections_1		0.1004	-0.0069	0.1170	-0.0112	0.0050
[p value]		0.1193	0.9918	0.0956	0.9826	0.9951
sd.injections_5		0.0647	-0.0052	0.0061	0.0003	0.0027
[p value]		0.5001	0.9948	0.9902	0.9997	0.9979
AIC (MNL, full sample)		44554.36	16600.5	32191.09	24083.88	24457.24
AIC (MXL, full sample)		42785.54	16341.2	31362.32	23848.89	24364.66

AIC Akaike Information Criterion, DM decision making, MNL multinomial logit, MXL mixed-logit, sd.XXXX standard deviation of random parameter estimates

Estimates of WTP, including the ‘option value’ of treatment, excluding China because of insignificant preferences

over cost among this sample, are shown in Table 6 and illustrated in Fig. 4. A WTP greater than zero indicates a WTP

Fig. 1 Part-worth utilities and confidence intervals by attribute and country/region



to *secure* a change from the reference level, whilst values less than zero indicate a willingness to pay to *avoid* a move from the reference level. [28] These results imply a substantial option value of treatment, with a range from €11,000 in the Nordic countries to more than €28,000 in USA. In terms of treatment attributes, the greatest WTP was associated with improved effectiveness (likelihood of live birth), where respondents from the Nordic countries, Spain and the UK were willing to pay between €3000 and €3500 for a 15% absolute improvement in effectiveness (or to avoid a 15% absolute reduction in effectiveness), whilst respondents from the USA were willing to pay more than €6000. Respondents from the Nordic countries and the USA were also willing to pay up to €3000 to move from ‘some’ to ‘full’ shared decision making, but less willing to pay to move from ‘no’ to ‘some’ shared decision making. There was also a WTP for a greater degree of shared decision making, and a reduction in treatment discomfort, although significant heterogeneity means that this latter WTP was not significantly different from zero in all samples. Respondents from the USA had a substantially greater WTP than other respondents for the option of treatment and for gains in effectiveness but

were similar to other respondents with respect to WTP for changes in other attributes.

As noted above, fast and inconsistent respondents showed significant divergence from other respondents. However, excluding the 861 respondents (11.3% of all respondents) who completed the DCE tasks in less than half of the median time for their country and who were inconsistent in the repeated task improved Akaike Information Criterion in all models but did not substantively alter the pattern or magnitude of preferences or estimates of WTP, including the insignificant preference over cost in the Chinese sample. The results of this secondary analysis are available in the ESM.

4 Discussion

We find that the effectiveness of treatment was the most (or second most) important attribute across all samples. However, notwithstanding its relative importance and substantial WTP in all regions, we do not see a dominant preference for effectiveness to the exclusion of other considerations, or a wide gap between effectiveness and other attributes in terms

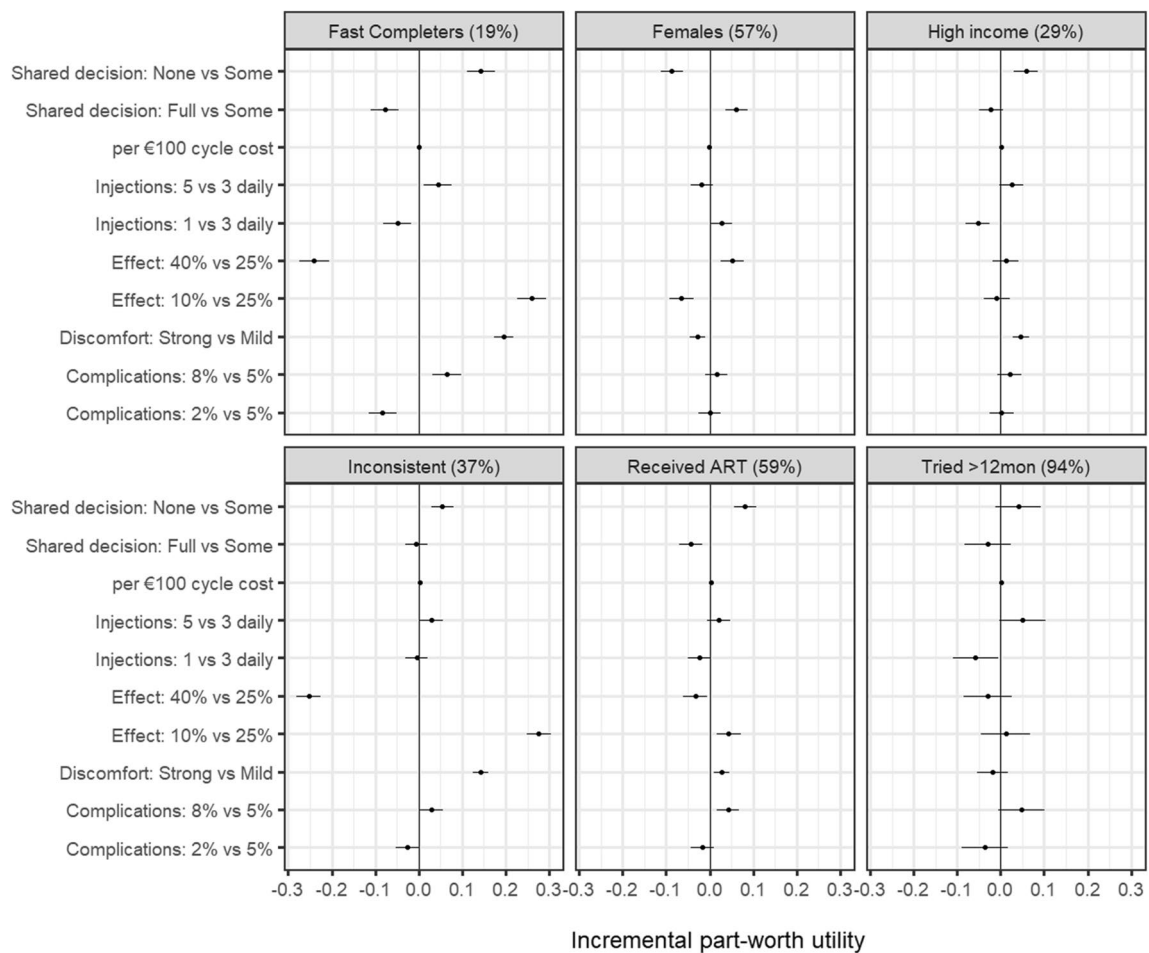


Fig. 2 Preference heterogeneity by respondent subgroups (subgroup as proportion of all respondents). *Fast Completers* completed the discrete choice experiment in less than half the country-specific median completion time (19% of respondents) relative ‘non-fast’ respondents. *ART* assistive reproductive therapies, *Females* female individuals (57% of respondents) relative to all other respondents (including “no answer”), *High Income* income quintiles 4 and 5 (29% of respondents)

relative to quintiles 1–3, *Inconsistent* chose a different alternative in the repeated task (37% of respondents) relative to those who were consistent in their choice, *No LT relationship* not in a long-term relationship (7% of respondents) relative to those in a long-term relationship or married, *Received ART* previously received medical assistance (59% of respondents) relative to those who did not receive assistance

of relative importance. Rather, we see evidence of simultaneous consideration of cost (in most regions), as well as aspects such as the (dis)comfort of treatment and the degree of shared decision making in treatment (Fig. 4).

Respondents placed a significant value on access to treatment, reflected in the ‘option value’ of treatment, but also had a substantial WTP for improvements in the effectiveness of treatment, a greater degree of shared decision making, and among some respondents, less discomfort in treatment. In general, respondents from the USA had the greatest WTP for improvements in effectiveness and, along with respondents from the Nordic countries, a greater degree of shared decision making. Respondents from other countries had a similar pattern but typically a lower absolute WTP.

The insensitivity of Chinese respondents to the cost of treatment led to a small and insignificant coefficient on the cost attribute and it was not appropriate to use it in WTP calculations. However, we observe that preferences over other attributes in the Chinese sample were broadly consistent with expectations. As in the other samples, we see that effectiveness was quite important, whilst the number of daily injections was relatively unimportant. Excluding potentially inattentive respondents did not change this pattern of preferences or the price insensitivity in this sample.

We have no data to explain why Chinese respondents were price insensitive over the range of cost presented. Given that the Chinese sample had the lowest proportion of respondents flagged as ‘fast and inconsistent’, we do not believe that this unexpected result is driven by a high degree

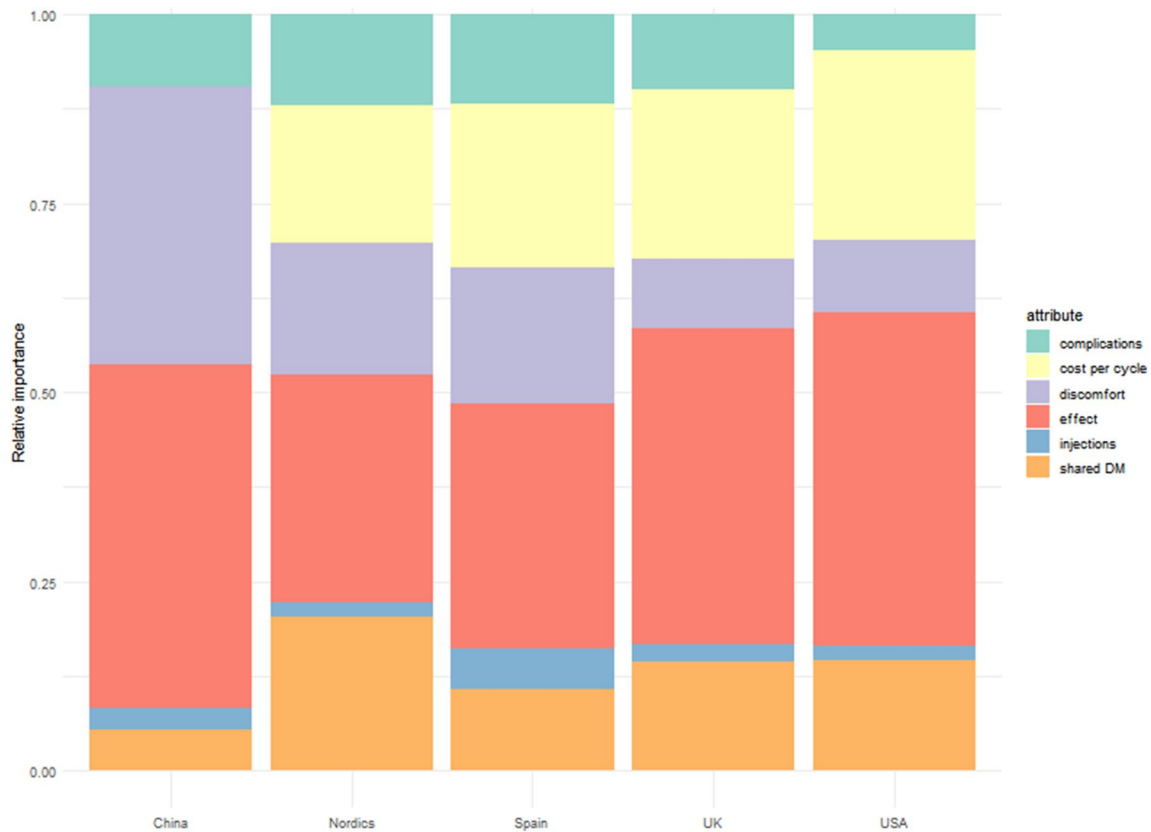


Fig. 3 Attribute relative importance by country/region. Values show attribute relative share of change in aggregate utility from least preferred to most preferred scenario. *DM* decision making

of inattentive respondents. In addition, outside of low price sensitivity, there was no obvious evidence of random or irrational responses, such as a high proportion of statistically insignificant attributes or ‘objectively irrational’ preferences such as a preference for lower effectiveness treatments. An alternative hypothesis is that the range of costs presented to the Chinese sample was too narrow for respondents to form a significant preference over the range of values and therefore respondents disregarded it in their choices. We used the same proportional range ($\pm 40\%$) that was associated with statistically significant price-sensitivity preference in the other samples, but if the midpoint estimate was artificially low for China (perhaps as a result of subsidised treatment costs), then, this range may be inappropriate. Finally, this result may reflect some ‘social desirability’ bias (SDB), whereby respondents may have felt pressured by cultural expectations to prioritise children over wealth, leading to price insensitivity in their (hypothetical) responses. ‘Social desirability’ bias may be particularly likely in the context of emotive topics such as parenting and infertility.

The hypothesis of ‘cultural’ price insensitivity driven by a greater degree of SDB in the Chinese sample is consistent

with results from the societal WTP portion of this survey [16], where Chinese public respondents had the highest stated maximum WTP for a national ART programme by a substantial margin. ‘Social desirability’ bias could also explain the substantial treatment-specific constants, or ‘option value’, in the Chinese sample, as consistently choosing *any* treatment over no treatment, regardless of the attribute levels of those treatments, would inflate ‘option value’. These results are consistent with the findings of a recent study that demonstrated SDB was associated with inflated valuations in an environmental economic study [33], and other research found that SDB was relatively stronger in more “collectivistic” countries [34].

To the extent that the price insensitivity of the Chinese sample reflects some degree of SDB, a key limitation to our study was not anticipating and controlling for this possibility. We do not, however, see substantive differences between the preferences of the Chinese sample and the other respondents in the other attributes. This suggests that although SDB may have led many Chinese respondents to disregard the cost attribute, this bias does not appear to have carried over to the other attributes. Indeed, the high

Table 6 Willingness to pay and 95% confidence intervals by attribute level and country/region (Euro)

Attribute	Change	Nordic countries	Spain	UK	USA	China
'Option value' ^b	Treatment vs no treatment	11,098.35^a 10,095.56; 12,101.14	13,615.13^a 12,967.91; 14,262.34	12,000.61^a 11,347.04; 12,654.18	28,217.86^a 26,556.80; 29,878.91	As preferences over cost were insignificant in the Chinese sample, meaningful estimates of willingness to pay cannot be inferred from the part-worth utilities
Likelihood of success	25% to 10%	-3340.9^a -1399.30; -5335.70	-3238.95^a -1564.12; -5200.95	-3556.84 -1841.67; -5308.37	-6161.88^a -4938.08; -14,233.36	
	25% to 40%	3575.74^a 6401.52; 836.86	3422.58^a 5598.78; 1,255.24	3399.00^a 6081.04; 725.96	6290.66^a 16,305.13; 1946.53	
Risk of complications	5% to 2%	1224.16^a 1662.73; 785.58	1030.57^a 1907.29; 203.53	784.42^a 1059.51; 509.32	662.25 1365.32; -40.81	
	5% to 8%	-1540.63^a -1108.34; -1972.91	-1359.61^a -1088.06; -1631.16	-824.24^a -554.51; -1093.97	-1308.61^a -606.34; -2010.89	
Discomfort	Strong to mild	-4625.24^a -18,159.16; 4860.13	-4701.54^a -19,374.17; 5957.98	-1615.53^a -8595.87; 3829.59	-4275.56 -23,048.14; 10,268.30	
Shared decision making	Some to none	-1810.23^a -1381.69; -2238.77	-753.71^a -485.55; -1021.88	-779.36^a -512.76; -1045.96	-1798.31^a -1102.88; -2493.75	
	Some to full	2820.18 5803.56; -11.16	1422.61 2948.50; -54.07	1592.11^a 2887.80; 248.13	3041.30^a 7743.07; 665.32	
Daily injections	3 to 1	430.59* 854.94; 6.24	636.63^a 903.46; 369.81	190.02 458.17; -78.14	514.95 1213.76; -183.87	
	3 to 5	-22.83 401.54; -447.2	-469.1^a -202.95; -735.24	-175.83 91.03; -442.7	-302.61 392.13; -997.34	

A positive value indicates that respondents would theoretically be willing to pay to secure a move to a more preferred level and a negative value indicates that respondents would theoretically be willing to pay to avoid a move to a less preferred level

^aSignificant at a 95% confidence level

^bOption value averages across treatment-specific constants A and B

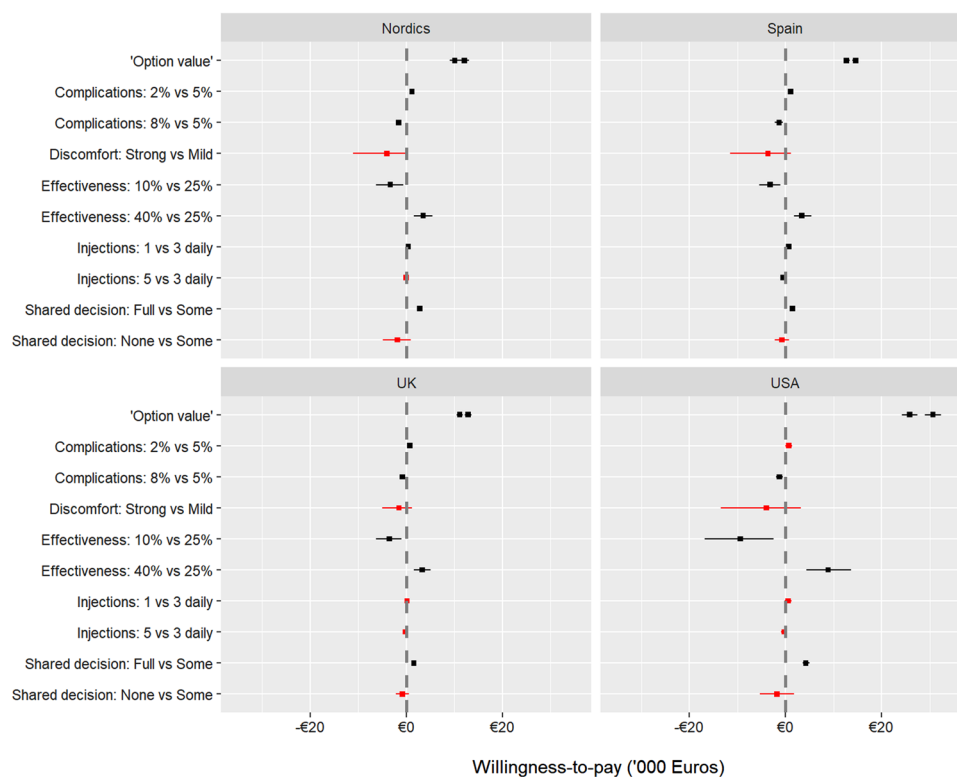
importance of discomfort in the Chinese sample indicates that respondents were willing to trade-off some chance of conception for a less uncomfortable treatment, suggesting their decisions were not driven by a sense of 'conception at any cost'. There is evidence that anonymous online surveys can reduce SDB [35], but future research should seek more effective methods to mitigate against this bias.

Another potential limitation is that respondents could have interpreted the wording of "no improved chance of pregnancy" with no treatment to mean "no chance of conception". This could have unintentionally encouraged some respondents to choose one of the treatment options over no treatment, thus inflating the option value of treatment. However, as likelihood of spontaneous conception after more than 12 months of trying is typically less than 10% [36], the practical difference between "no change" and "no chance"

is relatively small, for 90% of people, the outcome of no treatment will be the same (no pregnancy). For this reason, we believe that any bias in our estimates of optional value from this wording will be minimal.

Putting the overall results into the context of the existing literature, the relatively low importance of the number of injections and the high importance of the likelihood of success and of cost observed here are consistent with Musters et al. [10], who reported that an additional daily injection did not alter women's treatment preferences but that they were impacted by cost and the live birth rate. This study was conducted in the Netherlands and included 206 respondents. The authors found that, on average, respondents were only willing to pay €1000 if it was associated with an improvement in the live birth rate of at least 6%. Similarly, Palumbo et al. [9] reported a willingness to pay of between €100 and

Fig. 4 Willingness to pay by attribute and region, excluding China. Confidence intervals shown in *red* cross zero and are considered statistically insignificant



€300 for a 1–2% improvement in effectiveness. They also found that positive doctor-patient information sharing was more important to patients than treatment comfort.

van Empel et al. [6] specifically tested the impact of ‘patient centredness’ on patient and physician preferences for fertility care. They did not include a cost attribute but asked respondents to trade-off between the pregnancy rate and process aspects of treatment, including travel time to the clinic, the physician’s attitude toward the patient, the information provided to the patient and the continuity of care. They found that patients were willing to accept up to a 10% lower pregnancy rate for a friendly and interested physician, and for clear and customised information on treatment. Physicians given the same tasks and asked to anticipate patient responses underestimated the value of clear and customised information by more than 40% (a 5.5% trade-off in the pregnancy rate compared with the patients’ 9.6%). Our results are consistent with this finding, but we see a non-linear value to shared decision making: moving from ‘none’ to ‘some’ shared decision making was considerably less valuable than moving from ‘some’ to ‘full’ in all samples. Shared decision making, though, is a difficult concept to quantify, particularly compared with some of the other attributes in the DCE. Different participants may have had varying perceptions of what represented an acceptable degree of shared decision making in this context, and future research should seek to understand which ART decisions patients are most interested in sharing and which they prefer to delegate to their physician.

5 Conclusions

This study provides evidence from large multi-national samples to generalise the results of previous smaller scale research around patient preferences for ART. We find that the direction of preferences over attribute levels is relatively uniform across the countries/regions in the sample, but that the relative importance of those attributes can differ substantially. We also see that respondents balanced concerns for treatment effectiveness with other considerations, including the cost and (dis)comfort of treatment, and the degree of shared decision making. Moreover, we find a substantial ‘option value’ to treatment, demonstrating the value of access to ART to those with experience of subfertility.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40271-021-00563-7>.

Acknowledgements The authors gratefully acknowledge the contributions of Gayathri Kumar of the Office of Health Economics in preparing the manuscript.

Declarations

Funding This work was supported by a research contract from Ferring Pharmaceuticals. Ferring Pharmaceuticals approved the decision to submit the manuscript.

Conflicts of interest/competing interests Chris Skedgel is an employee of Office of Health Economics, which received a research contract from Ferring Pharmaceuticals for the conduct of this study and has received research grants from Ferring Pharmaceuticals for other projects in infertility. Eleanor Ralphs is an employee of IQVIA, which received a research contract from Ferring Pharmaceuticals for the conduct of this study as well as research grants/contracts from other pharmaceutical clients. Elaine Finn is an employee of IQVIA, which received a research contract from Ferring Pharmaceuticals for the conduct of this study as well as research grants/contracts from other pharmaceutical clients. Marie Markert is an employee of Ferring Pharmaceuticals. Carl Samuelson was a contractor to Ferring Pharmaceuticals at the time of the study and received personal fees in connection to his involvement in the project. Jennifer A. Whitty received personal fees from Ferring Pharmaceuticals in connection to the conduct of this study and has received personal fees from other pharmaceutical companies in the past.

Ethics approval The PIS, questionnaire, and the statistical analysis plan were reviewed and approved by the University of East Anglia Faculty of Medicine and Health Science Ethics Committee, Norwich UK (reference 201819-090).

Consent to participate All participants were required to indicate their voluntary consent to participate on the PIS before they were allowed to proceed to the questionnaire.

Consent for publication All authors give the publisher permission to publish this work.

Availability of data and material Data are available from the corresponding author upon reasonable request, subject to approval by Ferring Pharmaceuticals.

Code availability The statistical code may be shared upon reasonable request to the corresponding author.

Author contributions This work was a joint effort. All authors (CSk, ER, EF, MM, CSa and JW) made substantial contributions to the conception of the study, and the design and wording of the questionnaire. CSk, ER, EF and MM developed the estimates of the cost of ART in each country. CSk, ER and EF contributed to the data analysis and visualisation. All authors (CSk, ER, EF, MM, CSa and JW) contributed to the interpretation of the data. CSk, ER and EF produced the first draft of the manuscript, and JW, MM and CSa made substantial intellectual contributions to the final draft. All authors (CSk, ER, EF, MM, CSa and JW) provided their approval for submission and agreed to be named on the publication. All authors (CSk, ER, EF, MM, CSa and JW) agree to be accountable for all aspects of the work, and to ensure that any questions related to the accuracy and integrity of any part of the work are appropriately investigated and resolved.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

- Pablos-Mendez A, Radloff SR, Khajavi K, Dunst SA. The demographic stretch of the arc of life: social and cultural changes that follow the demographic transition. *Glob Health Sci Pract.* 2015;3:341–51.
- Kirk D. Demographic transition theory. *Popul Stud.* 1996;50:361–87.
- Virtanen HE, Jørgensen N, Toppari J. Semen quality in the 21st century. *Nat Rev Urol.* 2017;14:120–30.
- Ledger WL. Demographics of infertility. *Reprod Biomed Online.* 2009;18(Suppl. 2):11–4.
- Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum Reprod.* 2007;22:1506–12.
- van Empel IWH, Dancet EAF, Koolman XHE, Nelen WLDM, Stolk EA, Sermeus W, et al. Physicians underestimate the importance of patient-centredness to patients: a discrete choice experiment in fertility care. *Hum Reprod.* 2011;26:584–93.
- Cai QF, Wan F, Dong XY, Liao XH, Zheng J, Wang R, et al. Fertility clinicians and infertile patients in China have different preferences in fertility care. *Hum Reprod.* 2014;29:712–9.
- Gonen LD. Satisfaction with in vitro fertilization treatment: patients' experiences and professionals' perceptions. *Fertil Res Pract.* 2016;2:6.
- Palumbo A, De La Fuente P, Rodríguez M, Sánchez F, Martínez-Salazar J, Muñoz M, et al. Willingness to pay and conjoint analysis to determine women's preferences for ovarian stimulating hormones in the treatment of infertility in Spain. *Hum Reprod.* 2011;26:1790–8.
- Musters AM, de Bekker-Grob EW, Mochtar MH, van der Veen F, van Mello NM. Women's perspectives regarding subcutaneous injections, costs and live birth rates in IVF. *Hum Reprod.* 2011;26:2425–31.
- Botha W, Donnelly N, Shanahan M, Norman RJ, Chambers GM. Societal preferences for fertility treatment in Australia: a stated preference discrete choice experiment. *J Med Econ.* 2019;22:95–107.
- Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics.* 2019;37:201–26.
- Berg Brigham K, Cadier B, Chevreul K. The diversity of regulation and public financing of IVF in Europe and its impact on utilization. *Hum Reprod Oxf Engl.* 2013;28:666–75.
- Human Fertilisation & Embryology Authority. Fertility treatment 2017: trends and figures. 2019. <https://www.hfea.gov.uk/media/2894/fertility-treatment-2017-trends-and-figures-may-2019.pdf>. Accessed 17 Nov 2021.
- RESOLVE. Discover infertility treatment coverage by U.S. State. The National Infertility Association. 2020. <https://resolve.org/what-are-my-options/insurance-coverage/infertility-coverage-state/>. Accessed 19 Oct 2020.
- Skedgel C, Ralphs E, Finn E, Whitty JA, Markert M, Samuelson C. Is the public supportive and willing to pay for a national assistive reproductive therapies programme? Results from a multicountry survey. *BMJ Open.* 2021;11. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7949370/>. Accessed 27 Apr 2021.
- Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics.* 2014;32:883–902.
- Poder T, He J, Simard C, Pasquier J-C. Willingness to pay for ovulation induction treatment in case of WHO II anovulation: a study using the contingent valuation method. *Patient Prefer Adherence.* 2014;8:1337–46.

19. van den Wijngaard L, van Wely M, Dancet EAF, van Mello NM, Koks CAM, van der Veen F, et al. Patients' preferences for gonadotrophin-releasing hormone analogs in in vitro fertilization. *Gynecol Obstet Invest.* 2014;78:16–21.
20. IVF Worldwide. The costs of IVF in different countries. IVF Worldw. 2008. <https://ivf-worldwide.com/education/introduction/ivf-costs-worldwide/the-costs-of-ivf-in-different-countries.html>. Accessed 30 Mar 2020.
21. Collins JA. An international survey of the health economics of IVF and ICSI. *Hum Reprod Update.* 2002;8:265–77.
22. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? *Expert Rev Pharmacoecon Outcomes Res.* 2017;17:531–42.
23. Scott A. Identifying and analysing dominant preferences in discrete choice experiments: an application in health care. *J Econ Psychol.* 2002;23:383–98.
24. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Econ.* 2006;15:797–811.
25. Hensher DA, Rose JM, Greene WH. *Applied choice analysis: a primer.* Cambridge: Cambridge University Press; 2005.
26. Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Econ.* 2007;16:827–40.
27. Gonzalez JM. A guide to measuring and interpreting attribute importance. *Patient.* 2019;12:287–95.
28. Small KA, Rosen HS. Applied welfare economics with discrete choice models. *Econometrica.* 1981;49:105–30.
29. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. <https://www.R-project.org/>. Accessed 17 Nov 2021].
30. Croissant Y. Estimation of random utility models in R: the mlogit package. *J Stat Softw.* 2020;95:1–41.
31. Wickham H. *ggplot2: elegant graphics for data analysis.* New York (NY): Springer-Verlag; 2016. <https://ggplot2.tidyverse.org>. Accessed 17 Nov 2021.
32. Kassambara A. *ggpubr: "ggplot2" based publication ready plots.* 2020. <https://CRAN.R-project.org/package=ggpubr>. Accessed 17 Nov 2021.
33. López-Becerra EI, Alcon F. Social desirability bias in the environmental economic valuation: an inferred valuation approach. *Ecol Econ.* 2021;184:106988.
34. Kim SH, Kim S. National culture and social desirability bias in measuring public service motivation. *Adm Soc.* 2016;48:444–76.
35. Leggett CG, Kleckner NS, Boyle KJ, Duffield JW, Mitchell RC. Social desirability bias in contingent valuation surveys administered through in-person interviews. *Land Econ.* 2003;79:561–75.
36. Taylor A. Extent of the problem. *BMJ.* 2003;327:434–6.