Patient preferences for characteristics differentiating ovarian stimulation treatments

Erik Landfeldt¹, Barbara Jablonowska², Elisabeth Norlander³, Karin Persdotter-Eberg⁴, Ann Thurin-Kjellberg⁵, Margaretha Wramsby⁶, and Oskar Ström⁷,*

¹OptumInsight, Stockholm, Sweden ²Reproduktionsmedicinskt Centrum (RMC) Universitetssjukhuset, Linköping, Sweden ³Merck Serono AB (an affiliate of Merck KGaA, Darmstadt, Germany), Stockholm, Sweden ⁴Karolinska Universitetssjukhuset Huddinge, Stockholm, Sweden ⁵Sahlgrenska Academy, Gothenburg University, Göteborg, Sweden ⁶IVF-kliniken Cura, Malmö, Sweden ⁷Department of Learning, Informatics, Management and Ethics, Medical Management Centre, Karolinska Institutet, Berzelius väg 3, Stockholm SE-171 77, Sweden

*Correspondence address. Tel: +46-70-432-27-17; E-mail: oskar.strom@ki.se

Submitted on August 16, 2011; resubmitted on November 28, 2011; accepted on December 7, 2011

BACKGROUND: Little is known concerning patient preferences for IVF treatments. The objective of this study was to elicit patient preferences for characteristics differentiating ovarian stimulation treatments.

METHODS: Women undergoing IVF were recruited from six clinics in Sweden between May 2010 and December 2010. Included patients completed a study questionnaire consisting of one contingent valuation (CV) question (with six different bids) and 16 conjoint analysis (CA) questions formulated as discrete choices between two hypothetical ovarian stimulation treatments (defined in terms of manufacturing method, method of administration, time required for administration, dose variability and hypothetical price). Patient preferences were derived using multinomial logit modelling.

RESULTS: The final study population consisted of 294 women (mean age of 35). Respondents were willing to pay €360 [95% confidence interval (CI): €340–€390] to receive FSH derived from DNA technology instead of highly purified extract from urine from post-menopausal women, €300 (95% CI: €280–€320) to administer the FSH using a prefilled injection pen instead of a conventional syringe, €30 (95% CI: €20–€40) per saved minute required for administration and €530 (95% CI: €500–€570) to reduce the dose variability from 10–20% to 1–2% (P < 0.001 for all estimates). The result from the CV was similar to the CA.

CONCLUSIONS: Women undergoing IVF place significant value on characteristics differentiating ovarian stimulation treatments. Product-specific aspects should be taken into account by decision-makers when discriminating between commercial gonadotrophins in clinical practice to align health-care decision-making with patient preferences and potentially improve the effectiveness of IVF interventions through enhanced patient satisfaction and treatment compliance. Preferences for treatment characteristics should also be considered in evaluations of ovarian stimulation products to capture their true value from a patient perspective.

Key words: IVF / FSH / contingent valuation / conjoint analysis / Sweden

Introduction

IVF is a well-established assisted reproduction technique for the treatment of infertility. Modern conventional IVF protocols involve ovarian stimulation by daily s.c. injections of FSH. Commercial gonadotrophins contain FSH extracted from urine from post-menopausal women (hMG) or developed through recombinant human DNA technology (r-hFSH). Various injection or reconstitution devices exist and the products are typically self-administered by the patients after careful instructions.

Randomized controlled trials have failed to identify any significant differences between hMG and r-hFSH with regard to live birth rate (Lehert et al., 2010). A recent review concluded that r-hFSH and urinary gonadotrophins are ‘equally effective and safe’ (van Wely et al., 2011). However, evidence of differences between hMG and r-hFSH concerning dose variability and purity exists (Howles, 2000; Ludwig et al., 2002; Hugues, Barlow et al., 2003; Bassett et al., 2009; Lehert et al., 2010). In addition, currently licensed products containing hMG or r-hFSH are differentiated by other characteristics, such as method of administration. Specifically, there are currently three
Evidence-based methods and patient preferences in IVF

Materials and Methods

Preference elicitation methodology

Ovarian stimulation treatments are differentiated by a variety of characteristics and patients value these characteristics differently depending on their preferences. However, an individual’s preferences are typically not known to decision-makers and must therefore be collected and estimated. There are in broad terms two methods for eliciting preferences: stated preference techniques and revealed preference techniques: the former is typically questionnaire based and relies on asking people hypothetical questions similar to a market research interview, while the latter techniques utilize evidence on how people behave in the face of real-world choices. Stated preference techniques were chosen for this study owing to the fact that patients in Sweden seldom explicitly choose between different treatment options during their IVF (and data are therefore lacking to inform revealed preference techniques).

The present study used two stated preference techniques for eliciting preferences and the WTP for characteristics differentiating ovarian stimulation treatments: CV and CA. Both methods are based on the assumption that a product, for example a treatment, can be described by its characteristics (e.g. administration method) and that patient preferences for a product are based on the levels (e.g. conventional syringe) of these characteristics.

CV is applied to directly estimate the monetary WTP for a product or condition, for example by giving respondents a choice to pay a certain predefined amount of money (referred to as a ‘bid’) to receive a specific product instead of another product (both products described in terms of their characteristics and levels). As an example, the respondent could be asked ‘Would you be willing to pay €500 to receive Drug A instead of Drug B during your treatment?’ The responses are then analyzed to estimate the mean WTP for Drug A versus Drug B.

In CA, preferences for product characteristics and levels are collected by asking individuals to choose between different hypothetical products. Patients are presented with several product alternatives and are instructed to mark the products they prefer. The responses are then analyzed using statistical models to elicit what characteristics and levels different respondents preferred. The method can be used to establish whether a given product characteristic is important, the relative importance of various characteristics and how individuals are willing to trade off between characteristics. The monetary WTP is estimated indirectly by including price (or a price proxy) as a characteristic. For additional information regarding CA and CV, the readers are referred to Ryan and Farrar (2000) and Klose (1999).

Previous research (Stevens et al., 2000) suggests that WTP estimates derived using CV and CA may differ. Both techniques were included in the present study to obtain both indirect and direct WTP estimates for currently licensed ovarian stimulation treatments and to contribute to the body of evidence concerning the comparativeness of CV and CA.

Patient population

Patients were recruited from three public clinics (Sahlgrenska Universitetssjukhuset, Göteborg; Reproduktionsmedicinskt centrum (RMC), Universitetssjukhuset, Linköping and Karolinska Universitetssjukhuset Huddinge, Stockholm) and three private clinics (Fertilitetscentrum, Carlanderska sjukhuset, Göteborg; IVF-kliniken Cura, Malmö and IVF-klinikens Öresund, Malmö) in Sweden between May 2010 and December 2010. Women were eligible for inclusion if they were undergoing IVF treatment, with or without history of previous IVF treatment, and willing and able to participate in the study. Patients were excluded if they were unable to read and write in Swedish. An estimated sample size of ~300 was deemed sufficient for the CA based on previous research (Orme, 2005).

Written consent was obtained from all patients prior to participation. Ethical approval was granted by the Regional Ethical Review Board in Stockholm. A pilot study with a total of 43 participants was conducted in 2009. The outcomes of the pilot study were used to make minor changes to the final version.

Questionnaire design

Each participant was asked to review and complete an online study questionnaire. The questionnaire consisted of six parts: (i) general information
about the study, (ii) information about currently licensed ovarian stimulation treatments and injection devices and their different characteristics (without disclosing brand names), (iii) a single CV question, (iv) 16 CA questions, (v) questions concerning the respondent (e.g. age, previous IVF experiences and income) and (vi) an open-ended free-text question where respondents were asked to state what they regarded to be the most important components of the IVF treatment. Cognitive debriefing interviews were not conducted.

The content of the questionnaire, as well as all information provided to the respondents throughout the study, was reviewed by two separate committees: (i) an expert panel consisting of IVF physicians and (ii) nurses from the participating clinics. The aim of the review was to ensure that the information provided to the respondents was objective, accurate, easily interpreted and not misleading. The clinicians reviewed the content of the information as well as the choice of words and phrasing.

The questionnaires were created using the computer software Sawtooth Software SSI Web (Sequim, Washington, USA). The CA questionnaire design (fixed orthogonal, non-randomized questions, characteristics and levels), discrete-choice scenarios (combinations of treatment characteristics and levels) and the number of questionnaire versions (n = 2) were generated in SSI Web to optimize overall design efficiency in terms of (i) minimal level overlap (each level is shown as few times as possible in each question), (ii) level balance (each level is shown approximately an equal number of times) and (iii) orthogonality (levels may be evaluated independently of other characteristics levels). The average product characteristic level efficiency was 0.92 (ideal design: efficiency = 1.00).

Selection of product characteristics and levels

Five characteristics, each consisting of two to five levels, were identified as the primary differentiating aspects of currently administered ovarian stimulation treatments. The characteristics and levels were chosen based on a non-systematic review of the literature, in collaboration with IVF physicians in an expert panel advisory board and group discussions with nurses from the participating clinics, and were revised based on the outcomes of the pilot study. The ovarian stimulation treatment was defined as a daily dose given for 10–16 days. All participants were clearly informed in writing that there was no evidence that any of the studied characteristics and/or levels would be associated with better effectiveness with respect to pregnancy rates (or any related outcome).

(1) Manufacturing method: commercial gonadotrophins contain FSH extracted from urine from post-menopausal women (hMG) or developed through r-hFSH. The two forms of FSH were included as levels of the manufacturing method characteristic.

(2) Method of administration: Three administration vehicles for s.c. ovarian stimulation were at the time of study available in Sweden: a disposable injection pen prefilled with premixed solution, a non-prefilled pen loaded with ampoules with premixed solution formulation and a conventional syringe where the drug (supplied in sterile vials as a lyophilized powder) is mixed and dissolved in sterile diluent prior to injection.

(3) Time required for administration: the amount of time required for the self-administration was included as a product characteristic with three levels: 3, 6 and 9 min.

(4) Dose variability: there are two commonly employed methods for assessing the gonadotrophin content of commercial ovarian stimulation treatments: the Steelman–Pohley in vivo rat bioassay (Steelman and Pohley, 1953) and optimized size exclusion high-performance liquid chromatography (SE-HPLC). The rat bioassay method has been shown to have a variability in measuring the content of gonadotrophin of up to 10–20% (Driebergen and Baer, 2003). SE-HPLC, only applicable to r-hFSH, has been reported to have estimated batch-to-batch variability as low as 1–2% (Driebergen and Baer, 2003) but evidence for higher variability exists (Bassett et al., 2009). In the present study, the extreme values of these two techniques, 1–2 and 10–20%, were included as levels of dose variability.

(5) Hypothetical price: price was included as a product characteristic to estimate the monetary WTP associated with the other characteristics. It was specified as an out-of-pocket amount of money (not reimbursed) that respondents were required to pay for the ovarian stimulation treatment, in addition to any other costs associated with the IVF procedure. Five levels were considered (Swedish kronor, SEK): SEK1000 (€110), SEK2500 (€280), SEK5000 (€560), SEK10 000 (€1110), SEK15 000 (€1670) and SEK20 000 (€2780).

The names and descriptions of the studied characteristics and levels were based on official patient information leaflets (when available) but modified to ensure that the text was easily interpreted, objective and not misleading. The names and descriptions were also reviewed by the expert panel and the nurses. The hypothetical price was specified in SEK and converted to euros (€) (rounded to nearest 10) using an exchange rate of €/SEK 0.110101.

The CV question

Participants were asked if they were willing to pay SEKX out-of-pocket to receive hypothetical Treatment A instead of hypothetical Treatment B (Table I). Patients were randomly assigned one of six different bids: SEK1000 (€110), SEK2500 (€280), SEK5000 (€560), SEK10 000 (€1110), SEK15 000 (€1670) or SEK20 000 (€2780). There was no ‘none’ option included. In line with previous research in the preference uncertainty literature (Sonia Akter et al., 2008), the questionnaire also contained a follow-up question where respondents were asked to rate how certain they were about their answer on a four-step scale (‘not certain at all’, ‘uncertain’, ‘certain’ and ‘definitely certain’).

CA questions

Participants were asked to complete a total of 16 CA questions. Each question was formulated as a discrete choice between two hypothetical ovarian stimulation treatments. An example of one of the hypothetical questions is presented in Table II. There was no ‘none’ option included.

### Table I The CV question (as it was presented in the study).

<table>
<thead>
<tr>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed using r-hFSH</td>
<td>Extracted from urine from post-menopausal women</td>
</tr>
<tr>
<td>Prefilled injection pen</td>
<td>Conventional syringe (break and mix ampoules)</td>
</tr>
<tr>
<td>Time required for administration (min)</td>
<td>3</td>
</tr>
</tbody>
</table>

Note: Swedish kronor (SEK). Patients were randomly assigned one of six different bids (SEKX): SEK1000 (€110), SEK2500 (€280), SEK5000 (€560), SEK10 000 (€1110), SEK15 000 (€1670) or SEK20 000 (€2780).
and were excluded from further analyses. Patients who failed to choose the treatment respondents were asked to choose between two treatments identical in all respects except price. Patients who failed to choose the treatment with the lowest price were considered to have irrational preferences and were excluded from further analysis. Additionally, 1 of the 16 questions was designed to test if respondents provided rational answers (rationality test), where respondents were asked to choose between two treatments identical in all respects except price. Patients who failed to choose the treatment with the lowest price were considered to have irrational preferences and were excluded from further analyses.

Two of the 16 questions were designed to identify inconsistent answers (consistency test), where patients were asked to choose between the same treatments at two different points in the questionnaire. Respondents who provided different answers for these two questions were excluded from further analysis. Additionally, 1 of the 16 questions was designed to test if respondents provided rational answers (rationality test), where respondents were asked to choose between two treatments identical in all respects except price. Patients who failed to choose the treatment with the lowest price were considered to have irrational preferences and were excluded from further analyses.

**Statistical analysis**

The answers from the CV question were analysed to elicit the monetary value that the respondents were willing to pay to receive hypothetical Treatment A instead of hypothetical Treatment B. The crude WTP was analysed using non-parametric statistical methods (Kriström, 1990). In addition, a logit model was estimated to analyse the association between the bid amounts and the proportion of accepted bids. The statistical analyses were executed in Stata 11 (StataCorp LP, Collage Station, TX, USA).

The answers from the 16 CA questions were analysed in Sawtooth Software SMRT (Sequim, WA, USA). A multinomial logit model was fitted to the choice data to estimate the average (aggregated) utility associated with each treatment characteristic level. It was assumed that ‘time required for administration’ and ‘hypothetical price’ were continuous variables linearly associated with utility in order to enable predictions of utility losses associated with any administration time and/or price. Only main effects (i.e. the independent impact of each product characteristic) were considered in the model estimation because of the absence of a priori theoretical justifications for interaction effects. The estimated utility associated with ‘hypothetical price’ was used to determine the marginal utility benefit from a one-unit price decrease. The marginal utility benefit was subsequently used to estimate the monetary value the respondents associated with a certain utility gain and, in consequence, specific treatments, treatment characteristics and characteristics levels.

**Subgroup analysis**

In Sweden, generally up to three IVF treatments are subsidized for couples who are involuntarily childless, provided that it can be considered medically meaningful and that the woman is below the age of 40 years (exceptions to these criteria exist in different regions of the country). With the subsidy, the couple pay a reimbursed amount of money for the ovarian stimulation treatment (maximum of €200 per year, irrespective of the full cost of the treatment) but do not bear the cost for the procedure. Women who are not eligible for the subsidy may choose to still undergo an IVF treatment and pay for the procedure themselves (in addition to the cost for the ovarian stimulation treatment at a maximum of €200 per year). Most publicly funded IVF treatments are conducted in public clinics, while non-subsidized treatments are usually performed in private clinics.

The analyses in the present study were primarily conducted on all respondents. In addition, given that there may be differences in preferences between patients conducting a subsidized IVF and those who pay for the procedure themselves (as well as potential differences in clinical practice between public and private institutions) respondents were also stratified by type of clinic (i.e. public/private). Moreover, the WTP to receive hypothetical Treatment A instead of hypothetical Treatment B (Table I) was also estimated in the CA for patients aged <35 years versus ≥35 years, with children versus respondents without children, of low social status [education: high school; occupation: unemployed, sick leave, other; monthly household income (before taxes): <€6110] versus high social status [education: college/university; occupation: employed, self-employed; monthly household income (before taxes): ≥€6110] and individuals who conducted their first IVF versus patients with experience of IVF procedures.

**Anchoring bias**

In the context of decision-making, anchoring is a cognitive bias that describes the human tendency to make decisions that are biased toward an initially presented value (Tversky and Kahneman, 1974). The present study included an investigation of anchoring bias by analysing if responders who were randomly assigned high CV bids also expressed a higher WTP in the CA, and vice versa.

**Results**

In total, 365 women undergoing IVF treatment participated in the study. Of these, 53 subjects did not answer all questions in the questionnaire, 11 individuals failed the rationality test and 7 patients did not provide consistent answers. The final study population consisted of 294 women. Table III contains a summary of descriptive characteristics of the participants, stratified by type of clinic (public/private).

**Contingent valuation**

In total, 65, 46, 46, 53, 36 and 48 women (n = 294) were asked to pay 110, 280, 560, 1110, 1670 and €2780 to receive hypothetical Treatment A instead of hypothetical Treatment B, respectively. The majority (55% and 161 subjects) accepted the bid, and 76% (122 out of 161 subjects) stated that they were ‘certain’ or ‘definitely certain’ about their answer. Fig. 1 shows the proportion of all women who accepted the bid, stratified by bid amount. Evident from the graph, 74, 70, 63, 38, 38 and 39% of all women accepted to pay 110, 280, 560, 1110, 1670 and 2780 to receive Treatment A instead of Treatment B, respectively. The corresponding proportions for women who stated that they were ‘certain’ or ‘definitely certain’ about their answer were 63, 61, 46, 25, 19 and 25%, respectively. The observed mean WTP (area under the curve) for all answers were €1460. For ‘certain’ and ‘definitely certain’ answers, the mean was €1080. Note, however, that the maximum WTP is

<table>
<thead>
<tr>
<th>Two ovarian stimulation treatments are presented below. If these were your only alternatives, which one would you prefer?</th>
<th>Treatment X</th>
<th>Treatment Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing method</td>
<td>Extracted from urine from post-menopausal women</td>
<td>Developed using rhFSH</td>
</tr>
<tr>
<td>Method of administration</td>
<td>Prefilled injection pen</td>
<td>Non-prefilled injection pen (loaded with ampoules)</td>
</tr>
<tr>
<td>Time required for administration (min)</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Dose variability (%)</td>
<td>10–20</td>
<td>1–2</td>
</tr>
<tr>
<td>Hypothetical price (£)</td>
<td>280</td>
<td>1670</td>
</tr>
</tbody>
</table>

**Table II** An example of one of the 16 hypothetical discrete-choice CA questions.
unknown since a non-trivial proportion (39% of all women and 25% of those who provided a ‘certain’ or ‘definitely certain’ answer) of all subjects were willing to pay the largest proposed amount to receive Treatment A instead of Treatment B.

The result from the logit model of the relationship between the different bids and the proportion of accepted bids (only ‘certain’ and ‘definitely certain’ answers, \( n = 215 \)) is shown in Fig. 2. The mean WTP (area under the curve) was calculated at €1450 [bootstrapped 95% confidence interval (CI95): €1020–3300]. The estimated coefficient for bid was statistically significant \( (P < 0.001) \).

**Conjoint analysis**

The results from the estimated multinomial logit model are presented in Table IV. The estimated utilities for ‘manufacturing method’, ‘method of administration’ and ‘dose variability’ are level-specific and zero-centred and should be interpreted within each characteristic (higher utility implies stronger preference, and vice versa). The relative characteristic importance reveals how respondents valued a specific treatment characteristic in relation to the other studied characteristics. The relative importance of ‘manufacturing method’ was on average 26%, ‘method of administration’, 22%, ‘time required for administration’, 14% and ‘dose variability’, 38%. ‘Hypothetical price’ was included as a product characteristic with the sole purpose to estimate the monetary WTP associated with other characteristics.

The monetary value associated with a certain utility gain was derived using the estimated disutility associated with a one-unit price increase. The estimated mean WTP for each product characteristic is presented in Table V. When the output from the CA was
utilized to reconstruct the dichotomous-choice CV question (Table I), the estimated WTP for hypothetical Treatment A versus hypothetical Treatment B was estimated at $\approx 1290 (CI_{95}: 1190–1410).

The results from the CA analysis were also utilized to estimate the WTP for hypothetical Treatment A versus hypothetical Treatment B (the CV question, Table I) for patients aged $\leq 35$ years (versus $\geq 35$), with children (versus without children), of low social status (versus high social status) and individuals who conducted their first IVF (versus patients with experience of previous IVF procedures).

The results were stable across the investigated patient strata, with a mean variation in the estimated WTP of 2% and a maximum deviation of 15% for patients with children compared with the estimate based on all respondents.

**Anchoring bias**

Respondents who were randomly assigned the two lowest CV bids ($110$ and $280$) expressed a lower WTP ($1040$) compared with

---

**Figure 1** Responses to the CV question (proportion of accepted bids, by bid). Note: the mean willingness to pay (WTP) was calculated as the area under the curve.

**Figure 2** Logit model of the responses to the CV question (proportion of accepted bids, by bid). Note: the black bars represent bootstrapped CI_{95}.

The mean WTP was calculated as the area under the curve.
women who were given bids of €560 and €1100 (WTP €1300). Respondents given the two highest bids (€1670 and €2780) expressed an even higher WTP (€1370).

**Discussion**

During the recent decade, the development of highly purified forms of commercial gonadotrophins has enabled ovarian stimulation by self-administered s.c. injections. In addition, several easy-to-use injection vehicles have been developed to facilitate self-administration. Although the efficacy and safety of currently licensed ovarian stimulation treatments is well documented, little is known about how patients value other treatment-specific characteristics, such as manufacturing method or time required for administration. Thus, given similar efficacy profiles (van Wely et al., 2011) but dissimilar product characteristics, it may be difficult for health-care practitioners to discriminate between available treatment options in clinical practice. The present study is to our knowledge the first attempt to elicit patient preferences for characteristics differentiating ovarian stimulation treatments.

**Patient preferences for ovarian stimulation treatments**

This study has shown that women undergoing IVF procedure have preferences for certain ovarian stimulation treatments and their characteristics. One of the more important characteristics investigated in the study was dose variability. In fact, respondents were prepared to pay €530 to reduce the dose variability from 10–20% to 1–2%. Responses to the open-ended free-text question suggest that possible explanations for this finding include fear of adverse events and ovarian hyperstimulation and that patients associate high-dose variability with poor ‘drug precision’ and overall ‘quality’.

In line with previous research (Pang et al., 2003; Platteau et al., 2003; Weiss, 2007), the respondents also expressed that they would prefer to administer the gonadotrophin using a prefilled or non-prefilled injection pen instead of a conventional syringe. The result that women prefer easy-to-use administration methods (i.e. convenience) comes as no surprise, especially in the light of previous research pointing out the physiological distress women may associate with the IVF procedure and that self-administration may be an intimidating exercise.

In the free-text question included in the questionnaire, patients also expressed a non-trivial WTP (€360) to receive FSH derived

<table>
<thead>
<tr>
<th>Table IV Utility associated with the studied levels and the relative importance of each characteristic.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utility</strong></td>
</tr>
<tr>
<td>Manufacturing method</td>
</tr>
<tr>
<td>Developed using r-hFSh</td>
</tr>
<tr>
<td>Extracted from urine from post-menopausal women</td>
</tr>
<tr>
<td>Method of administration</td>
</tr>
<tr>
<td>Prefilled injection pen</td>
</tr>
<tr>
<td>Non-prefilled injection pen (loaded with ampoules)</td>
</tr>
<tr>
<td>Conventional syringe (break and mix ampoules)</td>
</tr>
<tr>
<td>Time (min) required for administration</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>Dose variability (%)</td>
</tr>
<tr>
<td>1–2</td>
</tr>
<tr>
<td>10–20</td>
</tr>
<tr>
<td>Hypothetical price (€)*</td>
</tr>
<tr>
<td>110</td>
</tr>
<tr>
<td>280</td>
</tr>
<tr>
<td>1110</td>
</tr>
<tr>
<td>1670</td>
</tr>
<tr>
<td>2780</td>
</tr>
</tbody>
</table>

Note: Model $\chi^2 = 1848$, degrees of freedom = 4, P < 0.001 for all estimated utilities. The estimated utilities are level specific and zero centred and should be interpreted within each product characteristic (higher utility implies stronger preference, and vice versa).

*Time required for administration was assumed to be linearly associated with utility. The estimated coefficient was −0.04500.

*Price was assumed to be linearly associated with utility. The estimated coefficient was −0.15688.

<table>
<thead>
<tr>
<th>Table V WTP for the studied treatment characteristics levels.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WTP (€)</strong></td>
</tr>
<tr>
<td>Manufacturing method</td>
</tr>
<tr>
<td>DNA technology versus highly purified extract from urine from post-menopausal women</td>
</tr>
<tr>
<td>Method of administration</td>
</tr>
<tr>
<td>Prefilled injection pen versus conventional syringe</td>
</tr>
<tr>
<td>Non-prefilled injection pen versus conventional syringe</td>
</tr>
<tr>
<td>Prefilled injection pen versus non-prefilled injection pen</td>
</tr>
<tr>
<td>Time required for administration</td>
</tr>
<tr>
<td>Per saved minute</td>
</tr>
<tr>
<td>Dose variability</td>
</tr>
<tr>
<td>1–2 versus 10–20%</td>
</tr>
</tbody>
</table>

CI, confidence interval.
P < 0.001 for all estimated utilities.
using DNA technology instead of extracted from urine from postmenopausal women. Possible explanations for this result, derived from the responses to the open-ended free-text question, include preference for modern technology products, aversive attitudes concerning the use of products derived from urine from other women and that some respondents associate DNA technology with ‘quality’.

The least important characteristic included in the study was ‘time required for administration’. This is not surprising given that available treatments are expected to require relatively limited and similar time for administration. Moreover, several respondents explicitly stated that the time required for administration was subordinate to all other treatment characteristics and that they were willing to devote the time required for the ovarian stimulation.

Taken together, the results from this study show that patients do value many aspects of their IVF treatment, including characteristics not directly related to effectiveness (i.e. pregnancy) or safety (i.e. adverse events). This finding supports the outcome of an earlier CA applied to IVF in general (not specific treatments or products) by Ryan (1999). In addition, Ryan found that individuals in fact are willing to trade changes in the probability of delivering a child with other characteristics and that there is utility beyond health outcome.

**Respondents from public versus private clinics**

In Sweden, the vast majority of all subsidized IVF treatments are conducted in public clinics. Patients who are undergoing their IVF at a private clinic would consequently be expected to have a higher WTP since they are paying for the entire procedure themselves (i.e. they are already in a mindset of having to pay for the treatment). Indeed, respondents recruited at private clinics expressed a higher WTP for all studied characteristics compared with women enrolled from public clinics. Moreover, women from private clinics expressed greater emphasis on ‘manufacturing method’, ‘administration method’, ‘time required for administration’ and ‘dose variability’ compared with patients from public clinics. One interpretation of these results is that women from public clinics who receive the treatment essentially for free are slightly less picky about the characteristics of their ovarian stimulation product compared with those who have to pay for the IVF. In addition, a larger proportion of respondents from private clinics had previous experience of IVF treatments, and may as a result be more informed and thereby better at valuing different treatments and their characteristics compared with those doing their first procedure.

**Study design**

Although 15% of the respondents did not answer all questions in the questionnaire, there was evidence that women who successfully completed the survey understood the questions and answered them correctly. Specifically, only 7 respondents (2%) did not provide consistent answers and 11 patients (4%) failed the rationality test. Moreover, responses to the free-text question indicated that patients felt that they were adequately informed to answer the questions in the questionnaire, even though some women expressed that they mostly cared about results (i.e. pregnancy) and therefore had little interest in any of the studied treatment characteristics.

Empirical evidence suggests that WTP estimates derived using CV and CA often differ (Stevens et al., 2000). In contrast to these findings, the outcome of the CV analysis in this study was in fact quite similar to the CA results. The total deviation was <4% (27% for ‘certain’ and ‘definitely certain’ answers). It is not possible to determine whether this result is an outcome of chance or robust study design (e.g. well-formulated information to the respondents or easily interpreted questions). It should be noted, however, that the maximum WTP could not be identified in the CV analysis owing to the fact that quite a few respondents accepted the highest bid. The comparison between the CV and CA results should consequently be interpreted with some caution.

In line with previous research of decision-making in different disciplines (Adrian Furnham and Hua Chu Boo, 2011), we identified an anchoring bias driven by the bid in the CV question. Specifically, patients who received lower bids in the CV question expressed a lower WTP for hypothetical Treatment A versus hypothetical Treatment B in the CA, and vice versa. However, given that the distribution of respondents was even across the different CV bids (low bids, n = 111; middle bids, n = 99 and high bids: n = 84), the impact of the bias on the overall study results was minimal.

It should be noted that the present study did not include ‘effectiveness’ as a differentiating product characteristic, for two primary reasons. First, although there has been some controversy regarding the comparative efficacy of hMG and r-hFSH, a recent review concluded that the two gonadotrophins are equally effective (van Wely et al., 2011). Secondly, as indicated by the responses to the free-text question in the questionnaire, women who undergo IVF are predominantly concerned about their chances of having a baby. Consequently, including any efficacy-related outcome would have blurred the estimated preferences (and WTP) for all other product characteristics. In other words, although it is clear that women undergoing IVF place significant value on the studied treatment characteristics, their preferences would most likely diminish in relation to an outcome such as live birth rate. For these reasons, ‘effectiveness’ was not studied as a product characteristic (and could in fact not have been included given the objective to study the other differentiating characteristics).

**Implications for health-care policy**

Contemporary health economic evaluations are generally performed in an extra-welfarism paradigm. Compared with welfarism which aims at maximizing the well-being of the individual or the society, extra-welfarism is concerned with maximizing a defined ‘goal’. In the context of health and health care, the goal is typically a health outcome or an adjusted health outcome, such as quality-adjusted life-years. Thus, whereas the welfarism approach explicitly incorporates the preferences of patients (and the society), extra-welfarism does not (Bridges, 2003). This study has shown that women undergoing an IVF procedure have significant preferences for various characteristics of ovarian stimulation treatments. In the currently dominant extra-welfarism paradigm, these results are only relevant if the investigated characteristics have implications for health outcomes. Thus, assuming that, for example, ‘administration method’ influences compliance and that compliance influence effectiveness and/or safety, there may be an argument that decision-makers should pay attention to patient preferences for this particular product characteristic. In the welfarism
paradigm, on the other hand, the utility of patients are maximized based on their complete preference functions, and the estimated preferences for all treatment characteristics should consequently be taken into account in health economic evaluations and in clinical practice.

Limitations
A limitation of the present study concerns the sample size. Although previous research suggests that a sample of ~300 subjects is sufficient for CA, such a population may not be adequate for CV, in particular if respondents are stratified by many CV bids and levels of response certainty. The fact that a larger proportion of women accepted the highest bid (‘certain’ and ‘definitely certain’ answers) compared with the next highest bid (Fig. 1) is an indication that the sample size may have been insufficient for stratified analysis in this study. A related limitation concerns the maximum CV bid amount. A CV should ideally include bids that capture the total WTP of all responders, which was not the case in the present study since some women accepted the highest bid. However, answers from the open-ended free-text question in the questionnaire indicate that some women are prepared to pay ‘whatever it takes’ to receive the ‘best’ and most appropriate IVF treatment based on their clinical condition and needs. Thus, perhaps it would have been difficult to include a bid amount which all respondents would have declined. Furthermore, if the proportion of these women with very high WTP was evenly distributed across the different CV bids (which would be expected given a sufficient sample) the proportion of accepted bids would be expected to be similar in the higher end of the bid amount spectrum. Evident from Fig. 1, this was in fact the case in the analysis of all bids. We tried to alleviate the issue of an inadequate stratified CV sample and maximum bid amount by conducting a parametric analysis of the association between the bid amounts and the proportion of accepted bids. Another limitation with the CV question was that it assumed that patients preferred Treatment A over Treatment B. Thus, it is possible that some respondents who rejected the bid were indifferent or would have expressed a negative WTP (i.e. preferred Treatment B over Treatment A) if permitted.

A second limitation concerns the external validity of the study results. Subgroup analysis indicated that the estimates were fairly stable across different patient groups, which suggest that the discrepancy between the study population and the population of interest may be fairly limited, even if the sample is non-representative in terms of the included covariates. However, there may be other differences between the study sample and the population of interest that is not captured in the collected background variables and it is therefore not possible to draw any definite conclusions concerning the generalizability of the study results. Moreover, the fact that we included patients undergoing a subsidized IVF may have had implications for the external validity of the study (i.e. limit the possibility to generalize our findings to settings without a national reimbursement system). However, as the possibility to conduct an IVF in a private clinic at a greater cost is well known to patients in Sweden, the participating women should not have been completely unfamiliar with the idea of paying for their IVF treatment, even if their current treatment was subsidized. We tried to alleviate this potential problem further by specifying the hypothetical price as an out-of-pocket amount of money (not reimbursed) that respondents were required to pay for the ovarian stimulation treatment.

Lastly, there is some evidence that the CV ‘take-it-or-leave-it’ question format possibly underlies the danger of ‘yea-saying’ (Klose, 1999). The presence of this bias in the present study is unknown. However, given that we included a post-decisional question where respondents were asked to indicate the level of certainty about their ‘Yes/No’ decision and that we conducted all analyses for these certain responses separately, the problem of ‘yea-saying’ was expected to be limited.

Conclusions
Women undergoing IVF place significant value on characteristics differentiating ovarian stimulation treatments. Patients preferred low-dose variability, FSH derived from DNA technology over highly purified extract from urine from post-menopausal women, injection pens over conventional syringe and short administration times. Product-specific aspects should be taken into account by decision-makers when discriminating between commercial gonadotrophins in clinical practice to align health-care decision-making with patient preferences and potentially improve the effectiveness of IVF interventions through enhanced patient satisfaction and treatment compliance. Preferences for treatment characteristics should also be considered in evaluations of ovarian stimulation products to capture their true value from a patient perspective.

Acknowledgements
The authors gratefully acknowledge Leif Hägglund (IVF-kliniken Oresund, MEDEON Malmö, Malmö, Sweden) and Göran Westlander and Margareta Wood (Fertilitetscentrum, Carlanderska sjukhuset, Göteborg, Sweden) for their contribution to patient recruitment.

Authors’ roles
E.L. was involved in data management and analysis, and interpretation and writing the manuscript. B.J., K.P.E., A.T.K. and M.W. contributed to patient recruitment and manuscript review. E.N. was involved in study design and manuscript review. O.S. contributed to study conception and design and manuscript review.

Funding
This study was sponsored by Merck Serono AB, Stockholm, Sweden (an affiliate of Merck KGaA, Darmstadt, Germany).

Conflicts of interest
E.L. and O.S. have consulted for Merck Serono AB. E.N. is currently employed by Merck Serono AB. M.W. is currently a member of Merck Serono AB advisory board (fertility). A.T.K. was previously a member of Merck Serono AB advisory board (fertility).

References


