

12 March 2016

Benjamin M. Craig, PhD, Kim Rand-Hendriksen, PhD

EQ DCE Competition Description, Rules, and Procedures v1.1



Background & Aims

At the 2015 EQ Plenary, much discussion was devoted to the alternative analytic specifications for the interpretation of health preference data. Each specification entails three primary components: a model (e.g., linear regression), a cumulative density function (CDF; e.g., logit) and an estimator (e.g., maximum likelihood). The model component specifies the relationship between the parameters. The CDF links the model to the estimator. The estimator is a procedure that incorporates the data to identify the parameters of the model and their uncertainty. Analytic specification selection is a difficult discussion under optimal circumstances. Complexities related to differences in sampling, language, and database procedures (i.e., dropping eccentric respondents) compound its difficulty. From these discussions, many of the discussants realised that little progress is possible unless we find a common basis for specification comparison. **An idea emerged to run an exploratory study and to allow multiple teams to apply their own specifications independently (i.e., a head-to-head comparison).** For this purpose, Drs. Craig and Rand-Hendriksen developed and successfully submitted a proposal to the EuroQoL Research Foundation titled, “EQ DCE: Crowdsourcing innovation in valuation specification,” which was approved on 22 January 2016.

Health preference research (HPR) is inherently a scientific enterprise; specifications are devised, hypothesised, and tested. However, it is a conventional practice within EQ-5D valuation to conduct just one study and estimate just one analytic specification (i.e., no confirmation) to inform health technology assessments. Some studies include multiple specifications and “pick” among their results (i.e., data mining). Instead of comparing specifications based on their fit of the exploratory data, **the victor of this competition will be selected based on her or his predictions for a confirmatory study** collected after all specifications are made public.

Within HPR, the primary purpose of a health valuation study is to construct preference weights to be applied to real-world health outcomes (e.g., clinical trial endpoints). Although exploratory studies are constructed to efficiently identify the parameters (under an assumed specification), the primary objective is to predict the value of a broad range of outcomes. Therefore, **the confirmatory study will include EQ-5D-5L states commonly found in clinical and general population surveys.** This procedure will allow us to test the actual out-of-sample predictive ability of various specifications - arguably the ultimate test of their relative merits.

In summary, the competition entails: (1) an exploratory EQ DCE study that will allow multiple teams to apply their own specifications independently and (2) a confirmatory study that will include EQ-5D-5L responses commonly found in clinical and general population surveys. This head-to-head comparison will promote greater understanding of the merits underlying alternative specifications and may inform the design and analysis of future HPR studies.

Although the competition will have only one victor (as specified by the minimum chi-square), much can be learned regardless of the outcome. If a particular specification is found to clearly outpredict the rest, the conclusion would be to explore why and to also promote the exploration of this specification, not to mandate its practice. If all specifications are largely equivalent, the conclusion would motivate an investigation of how the most parsimonious model (the one with the least parameters) was able to do as well as the the more complex models.

Methods

The methods are best separated into 3 stages: **exploratory, prediction and confirmatory**. The exploratory stage involves (a) fielding of the exploratory study and (b) the invitation of HPR investigators to participate in the competition. The prediction stage entails (a) the registration of teams and abstract submission; (b) the receipt, audit and curation of submissions; and (c) the drafting of reports and working paper. The confirmatory stage involves (a) the fielding of the confirmatory study and presenting the results as well as (b) the submission of the final paper.

EXPLORATORY STAGE

Pair Selection: The pairs for the exploratory survey were based on those used in the original version of the Valuation Technology developed by the EuroQol Group (EQ-VT). As part of the EQ-VT, respondents

completed 7 paired comparisons (e.g., right). Each showed 2 EQ-5D states and respondents were asked “which is better, state A or state B?” Health states were described using EQ-5D-5L responses (See Table 1). Using these attributes and levels, Mark Oppe, a founding IAHP member and Senior Scientist at the EuroQol Group, selected 196 pairs for the EQ-VT.



Table 1 Adjectival Statements for EQ-5D-5L Attributes and Levels

| Attribute | # | Adjectival statement |
|--------------------|---|--|
| Mobility | 1 | No problems in walking about |
| | 2 | Slight problems in walking about |
| | 3 | Moderate problems in walking about |
| | 4 | Severe problems in walking about |
| | 5 | Unable to walk about |
| Self-care | 1 | No problems washing or dressing self |
| | 2 | Slight problems washing or dressing self |
| | 3 | Moderate problems washing or dressing self |
| | 4 | Severe problems washing or dressing self |
| | 5 | Unable to wash or dress self |
| Usual Activities | 1 | No problems doing usual activities |
| | 2 | Slight problems doing usual activities |
| | 3 | Moderate problems doing usual activities |
| | 4 | Severe problems doing usual activities |
| | 5 | Unable to do usual activities |
| Pain/Discomfort | 1 | No pain or discomfort |
| | 2 | Slight pain or discomfort |
| | 3 | Moderate pain or discomfort |
| | 4 | Severe pain or discomfort |
| | 5 | Extreme pain or discomfort |
| Anxiety/Depression | 1 | Not anxious or depressed |
| | 2 | Slightly anxious or depressed |
| | 3 | Moderately anxious or depressed |
| | 4 | Severely anxious or depressed |
| | 5 | Extremely anxious or depressed |

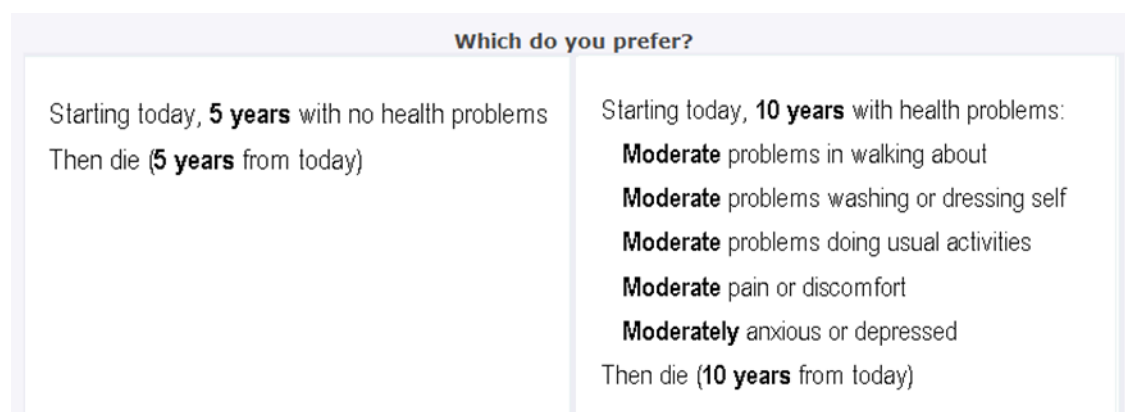
The exploratory pairs have two key differences compared to the original EQ-VT pairs. First, we constructed a superset of pairs that includes each of the 196 EQ-VT pairs and one of ten durations (1, 2,..., 10 years) for each alternative (196×10×10=19600). From this superset, Richard Norman selected the 600 pairs using NGENE (See Appendix A). We are grateful for his expertise and assistance with this selection process. His set of 600 only include 189 of the 196 original pairs; therefore, we include 10 examples of the 7 missing EQ-VT pairs. From this adjusted set of 610 pairs, we selected one pair for each of the 196 original pairs as well as 4 additional pairs that fill gaps in the level balance of the duration attribute. Among the 200 pairs, each combinations of durations ((10 choose 2) +10 = 45+10 = 55) appear 3 to 5 times. Each pair in the initial set of 200 pairs will be run with 4 different temporal units (10 years, 12 months, 10 weeks, and 30 days; Table 2), expanding the number of efficient pairs from 200 to 800.

The second difference compared to the original EQ-VT is that we changed the wording of the question and alternatives to facilitate health valuation on a QALY scale. We ask “Which do you prefer”, not “Which is better,” because survey is designed to elicit preferences, not judgments (e.g., voting). We dropped the words “A” and “B”, which may imply rank, and bolded the numbers and differentiating adjectives. We also explicitly describe the duration and timing of the health state (i.e., Starting today, [X] years with health problems: [health state] then die (X years from today)). In addition, this survey had different consent/directions and no interviewers.

Table 2. Adjectival Statements for Lifespan*

| # | Years | Months | Weeks | Days |
|----|-----------------|------------------|-----------------|-----------------|
| 0 | Immediate death | Immediate death | Immediate death | Immediate death |
| 1 | 1 year | 1 month | 1 week | 3 days |
| 2 | 2 years | 2 months | 2 weeks | 6 days |
| 3 | 3 years | 3 months | 3 weeks | 9 days |
| 4 | 4 years | 4 months | 4 weeks | 12 days |
| 5 | 5 years | 5 months | 5 weeks | 15 days |
| 6 | 6 years | 6 months | 6 weeks | 18 days |
| 7 | 7 years | 7 months | 7 weeks | 21 days |
| 8 | 8 years | 8 months | 8 weeks | 24 days |
| 9 | 9 years | 9 months | 9 weeks | 27 days |
| 10 | 10 years | 12 months | 10 weeks | 30 days |

*Highlights indicate lifespans of equal duration.



In complement to the efficient pairs, we include 800 time trade-off (TTO) pairs (see example below). A TTO pair includes one health state with problems and another without problems and are necessary for health valuation on a quality-adjusted life year (QALY) scale (below). These 800 pairs were selected using 3 steps:

1. Richard Norman predicted the equivalent lifespan with no health problems (i.e., QALY) for each health state in the 600 pairs. We divide these estimates by the duration (i.e., constant proportional assumption) to identify each state’s value on a QALY scale. We kept a candidate health state if its value was between 0.1 and 0.8 and it seemed probable (Jaccard index >0.005). Among the 392 states in the 196 EQ-VT pairs, 40 EQ-5D states met these 2 criteria.

2. The 40 health states were sorted by their value on a QALY scale and categorized into 4 groups of 10 (Extreme, Severe, Moderate and Slight). Each state within a group was paired with 5 durations with no health problems, creating 5 pairs (Extreme: 0,1,2,3,4; Severe: 1, 2, 3, 4, 5; Moderate: 3, 4, 5, 6, 7; Slight: 5, 6, 7, 8, 9).
3. Each of the 200 TTO pairs will be run with 4 different temporal units (10 years, 12 months, 10 weeks, and 30 days; Table 2), expanding the number of TTO pairs from 200 to 800.

In summary, the exploratory pairs include 800 efficient pairs and 800 TTO pairs (See superset and blank prediction files on IAHPR website).

Pair Selection for the Confirmatory Study: The confirmatory pairs were selected using the same process as the exploratory pairs, except that it is based on a superset of Jaccard states. A Jaccard state is a health description where the attributes are probabilistically similar. To identify which EQ-5D-5L states are Jaccard states, we estimate the Jaccard Index (i.e., $P(A \text{ and } B)/P(A \text{ or } B)$) for each pair of attributes using a dataset of 12,000 EQ-5D-5L responses. Given that each of the 3125 EQ-5D-5L states has 5 attributes, each has 10 Jaccard indices ($5 \text{ choose } 2 = 5 \times 4 / 2$). If all 10 indices were above 2% (i.e., threshold), the state was considered probable (140 of the 3125 states). Furthermore, we included an adjusted states for each Jaccard state with a maximum level 4. This adjusted state is that same as the original, except that we added one level. These adjusted states were added due to the sparse number of Jaccard states with level 5 on any domain. The confirmatory superset includes all non-dominant pairs of Jaccard states (see IAHPR website); however, we dropped those pairs where state A is worse than state B on only one domain and one level difference and A has an equal or longer lifespan than B. Using this superset (210585 pairs), Richard Norman ran NGENE to select 200 efficient pairs using the same program as the exploratory pairs. Each pair in the initial set of 200 confirmatory pairs will be run with 4 different temporal units (10 years, 12 months, 10 weeks, and 30 days; Table 2), expanding the number of efficient pairs from 200 to 800.

To aid the selection of the TTO pairs for the confirmatory survey, we examined its 200 efficient pairs. These pairs include 148 health descriptions and each description occurs 1 to 8 times. These pairs include 23 of the 392 descriptions in the exploratory pairs. The efficient confirmatory pairs include 10 durations and each duration occurs between 32 and 50 times. These pairs include 191 unique pairs of health descriptions (i.e., 9 appear twice), which is similar to the exploratory pairs (196 unique pairs of health descriptions and 4 appears twice). The 200 efficient pairs include only 47 of 55 potential differences in duration (i.e., unlike the exploratory survey) and these occur between 1 and 12 times.

Among the 148 health descriptions in the 200 efficient pairs, 67 have a predicted value on a QALY scale between 0.1 and 0.8. Among these 67 health descriptions, we dropped 7 descriptions, because they were in the exploratory survey. An additional 7 descriptions were dropped, because they are similar to descriptions in the timing, duration and lifespan study (i.e., $MO=SC=UA$ and $MO>1$). To achieve a final set of 40 descriptions across the range from 0.1 to 0.8, we had to arbitrarily drop 13 additional descriptions.

Each of the 40 descriptions were split into 4 groups and paired with 5 alternatives with no health problems, creating a initial set of 200 TTO pairs (identical to the exploratory TTO pairs). Each of these pairs will be run with 4 different temporal units (10 years, 12 months, 10 weeks, and 30 days; Table 2), expanding the number of TTO pairs from 200 to 800. In summary, the confirmatory pairs include 800 efficient pairs and 800 TTO pairs (See files on IAHPR website).

Statistical Power: The calculation of statistical power is complicated by the fact that the exploratory and confirmatory studies must have sufficient sample for a broad range of models. Each pair sample will have 50 or more respondents. At sample sizes of 50 and a type-1 error of 0.025, a 2-sided test will reject the equivalence of 2 pairs with contradicting population

probabilities (33% and 67%; minimum of 1 in 3 differences) with a power of 0.89. To achieve a normal approximation, the statistical rule of thumb (NP5) is that the product of the sample size, N_k , and the population probability, P_k , is greater than 5 (i.e., $N_k \times P_k > 5$ and $N_k \times (1 - P_k) > 5$). With 50 respondents, the sample probability, p_k , is approximately normally distributed if less than 90% of the population agrees (i.e., $50 \times P_k > 5$ and $50 \times (1 - P_k) > 5$). This sample size and population probability range place an upper bound on the standard errors of the sample probabilities ($\sqrt{0.5 \times (1 - 0.5) / 50} = 0.07$).

All sample probabilities, other than 0.5, will have an error < 0.07 due to their closer proximity to 0 and 1. If the population probability is outside of this range, the NP5 rule no longer holds. In this most extreme case, the study might incidentally include a pair in which all respondents agree ($p_k = 0$ or 1). To address this, the rule of 3 will be applied to estimate the confidence interval on the pair, similar to adverse events in clinical trials: if the sample is of size N_k , the 95% confidence interval is 0 to $3/N_k$ (e.g., 0 to 6% for $N_k = 50$). Assuming 50 responses per pair, a population probability between 0.1 and 0.9, and 20 responses per respondent, **the exploratory and confirmatory studies will each have 1600 pairs and 4000 respondents** ($1600 \times 50 / 20$).

Quota Sampling: To assure demographic representativeness of each pair probability, each pair sample will have the same quotas for gender, age and race/ethnicity categories (Table 3).

| Table 3. Pair-specific Quota Samples | Men | | | Women | | |
|--|-------|-------|-----|-------|-------|-----|
| | 18-34 | 35-54 | 55+ | 18-34 | 35-54 | 55+ |
| Prevalence according to 2010 US Census | | | | | | |
| White/Other, Non-Hispanic | 10% | 13% | 12% | 10% | 13% | 15% |
| Black, Non-Hispanic | 2% | 2% | 1% | 2% | 2% | 2% |
| Hispanic | 3% | 3% | 1% | 3% | 3% | 1% |
| Demographic Quotas for Each Pair Sample (N=50) | | | | | | |
| White/Other, Non-Hispanic | 5 | 7 | 7 | 5 | 7 | 7 |
| Black, Non-Hispanic | 1 | 1 | 1 | 1 | 1 | 1 |
| Hispanic | 1 | 1 | 1 | 1 | 1 | 1 |

Survey Instrument: For each study, the survey instrument will include 4 components: the screener, health, paired comparisons, and follow-up. The screener component will include a consent page and questions from the U.S. Census about respondent demographic characteristics, including sex and age. The health component will include the 5-level version of the EQ-5D, which will introduce the respondent to the EQ-5D descriptive system. The paired comparison component will include 3 examples: Apple/Orange, Good Health vs. Poor Health and Bad Health vs. Poor Health. The final example will serve to introduce the concept of subjectivity in the choice. After these examples, the respondents will complete 20 paired comparisons. Pairs will be randomly assigned at the survey level (no blocks). Each pair will be shown in a random sequence. The follow-up component will include an open comment question to allow respondents to express feedback on the survey instrument. Overall, the survey will have a median completion time of 25 minutes (from consent to submission).

The survey will also include 5 programs to support the survey instrument. (1) **Loading bar:** An animation that appears and rotates for multiple seconds before disappearing and allowing respondents to answer the questions on that page. It passively assures that the page fully loads and mandates that respondents take time to read the descriptions. (2) **Clickable image:** a utility that allows respondents to click the text that describes the health outcome instead of a button. Clicking the text causes the scenario to be enlarged and bordered in red, which may improve survey functioning on tablet devices. (3) **Matrix assignment:** A utility that automatically assigns a unique series of scenarios to respondents for their 30 paired comparisons and assures that each pair has a sufficient sample size ($N > 50$). (4) **Paradata capture:** A utility that automatically captures data about respondents' browsers, window sizes, and response times (e.g., changed responses). Paradata is nearly impossible to collect with paper-based surveys and improves the

assessment of data quality. (5) **IP checker:** A program that automatically identifies the geolocations of respondents' IP addresses (e.g., ZIP codes), and then de-identifies this information using hexadecimal encryption. Encryption protects respondents' anonymity and excludes respondents who attempt to take the survey twice or who use foreign or proxy servers.

Fielding the Exploratory Study: As with past surveys, all participants must: 1) be 18 years or older, 2) be able to read English and 3) reside in the United States. Also, persons who participated in the exploratory study will not be allowed to participate in the confirmatory study. The steps to a survey launch will include a hand-off meeting, redirect programming, testing and a soft launch. The hand-off meeting will include the research team and panel company and review the protocol and timeline. After the hand-off meeting, the panel company will construct and test their invitation system. At soft launch, 100 respondents will be sent a generic email invitation detailing payment information along with a survey link (no routing allowed). If successful, the team will request that the panel vendor recruit the remaining panelists. Each panelist who complete the 25-minute survey will be paid in points by the panel provider. Once the quota is filled, the data will be downloaded and audited for technical errors.

Analysis: The exploratory data will undergo an initial assessment for completeness and face validity at the aggregate level in order to verify proper coding. No respondents will be removed based on their responses to the paired comparisons; however, some variation is expected due to technical issues. The data will be completely de-identified prior to analysis.

PREDICTION STAGE

Registration of Teams: Each team must complete the Registration Form by 13 April 2016.

To register your team: <http://app.keysurvey.com/f/1018308/7ed4/>

This form is brief (only 5 questions): (1) Conditional Agreement for Teams; (2) Team name, team leader and number of co-investigators; (3) listing the co-investigators; (4) Experience with EQ-5D and DCE modeling; (5) IAHPR Sponsorship and Invoicing. Only one registration is allowed per team leader; however, leaders may participate as co-investigators on other teams.

For the Experience question (up to 500 words), each team leader will be asked to briefly describe her or his experience with EQ-5D and DCE modeling. If a team leader lacks experience, we recommend collaboration with researchers with greater experience with the EQ-5D and/or DCE modeling, particularly EuroQol Group or IAHPR members. Students within fields of relevance are welcome to compete. Drs. Craig and Rand-Hendriksen will review this description when selecting teams for the competition. This selection process is a necessary precaution to prevent proxy submissions and to assure that each submission has the potential to emerge victorious. Participants may be part of more than one team. However, Drs. Craig and Rand-Hendriksen retain the right to prevent attempts at "gaming" the registration process to maximize financial gain or chance of winning the competition.

During the registration of teams, we will be fielding the exploratory study, which should take about 4 weeks. Once a team has registered and been selected, each team will receive the same package of study materials (1 May), which includes:

1. An index of teams, including name and contact information of team leaders
2. Codebook and comma-separated values (csv) data file (7 columns [r_id, p_id, T_A, T_B, H_A, H_B, and choice] and 80000 rows [4000x20]). Namely, the dataset will include a unique respondent identifier (r_id; 1 to 4000), pair identifier (p_id; 1 to 1600), lifespan of the 2 alternative (T_A, T_B); vector descriptions of the 2 alternative states (H_A, H_B; e.g., 22232), and a binary choice identifier that represents preferring A (choice; 0 or 1)

3. Stata code that demonstrates the estimation of main effects under constant proportionality using (a) a linear probability model in ordinary least squares; (b) a logit model using maximum likelihood estimation and a rescaling parameter; and (c) a Bradley-Terry model using weighted least squares.
4. Prediction Submission Form and blank prediction file (5 columns[p_id, T_A, T_B, H_A, H_B] and 3200 rows [1600 exploratory pairs, 1600 confirmatory pairs])
5. Dr. Craig's entry for the Competition, including his form, prediction file and stata code.

A copy of a blank prediction file is currently available on the IAHPR website for reference.

Receipt, Audit and Curation of Submissions: Between 1 May to 4 July, each team will conduct their analyses and submit their findings. Basically, the deadline will be 2 weeks prior to the due date for the EQ Plenary papers. Each team must submit the following files:

1. Prediction Submission Form
2. Prediction file (6 columns[p_id, T_A, T_B, H_A, H_B, prediction] and 3200 rows [1600 exploratory pairs, 1600 confirmatory pairs])
3. Code and Log file

The **Prediction Submission Form** will capture details of their specifications, which will be incorporated into the working paper. In addition to their model and predictions, each team will provide a short description of the rationale and the procedure to select/create their specification. This motivation and history is crucial for the hypothesis generating stage. Each team will likely have a different approach to specification, particularly parameterisation.

This endeavor opens the floor to any approach to the analysis of the EQ DCE data - ranging from the standard 20-parameter logit model with maximum likelihood estimation to the use of a dousing wand, or printing the pairs on pieces of paper, throwing them to the wind, and measuring where they end up. The benefits of the competition is that it can maximize the creativity and innovation within a community of scientists, bring out a wealth of approaches.

The **prediction file** must be delivered as a csv file (6 columns[p_id, T_A, T_B, H_A, H_B, prediction] and 3200 rows [1600 exploratory pairs, 1600 confirmatory pairs]). The prediction file is identical to the blank prediction file (see IAHPR website) except with an additional column (prediction) that represents the team's predictions. Predictions are allowed up to 3 digits (e.g. 0.123), which is sufficient precision for pair samples of 50 respondents.

To characterise fit, we will compute the chi-square using the exploratory sample probabilities and predictions of each team :

$$\sum N_k \times (y_k - p_k)^2 \times (y_k \times (1 - y_k))^{-1}$$

In this formula, N_k is the sample size (i.e., 50 responses), p_k is the team's prediction, and y_k is the sample probability for the k^{th} pair. Although some teams may attempt complex models with multiple parameters (and achieve a smaller chi-square on the exploratory pairs), this may inadvertently overfit the exploratory dataset and poorly predict the confirmatory pairs. Teams may choose to relax the constant proportionality assumption as they see fit.

The **code and log files** must be delivered as a MS Word document (*.docx). We will rely on the honor system and will neither re-run the code nor double check that their log files are accurate, except to verify the victor of the competition. Unlike the Form and predictions, **the code and log file will NOT be posted without team approval**. We would be pleased to post a team's code on the IAHPR website, if they wish. Two Stata programs will be distributed with the exploratory dataset (example code and Dr. Craig's entry), which may facilitate the submission of other code.

Apart from chi-square computation, an audit for completeness, and copyedit for a language and format, the submissions will be largely unaltered. Each team that submits their results should

receive a notification on its completeness within 1 week of their receipt. Invoices are due with 14 days of approval notification. The forms for all complete submissions will be posted by 15 July (4th IAHPR Meeting).

Drafting of the Primary Paper: On 13 April, an abstract for the competition will be submitted to the 4th IAHPR Meeting and EuroQol Group Plenary, listing the number of teams and results of Dr. Craig's prediction entry. The two authors will be Drs. Craig and Rand-Hendriksen. Other teams may be interested in submitting their predictions as abstracts for the 5th IAHPR Meeting to be held on 2 September (prior to the EuroQol Plenary).

On 15 July, a working paper will be distributed include descriptions of the pair selection process, the survey instrument, the fielding, the exploratory results and the team results (teams, specifications and predictions). It will appear in the EuroQol Working paper series with 2 Appendices. Appendix 1 will include the pair results, namely the predictions of each team (by name) of the exploratory and confirmatory pairs. Appendix 2 will include the Prediction Submission forms from each team. The working paper will not include the parameters or code of any given specification, which should allow each team to submit its specification as a separate paper to a peer-reviewed journal independently (assuming the proper citation), if they see fit.

A concern that has been raised relates to intellectual property; will the teams be willing to make available their best ideas regarding EQ DCE modeling, and how do we maintain their claim to IP? First, we sincerely hope (and believe) that the teams will be willing to present their ideas. Each submission will be presented in full within the EQ Working Paper Series as part of Appendix 2. Thus, future work in which models or methods resulting from this project is used should refer to the original contributor, rather than the authors of the project manuscript (i.e. "We used the methods suggested by A and B in [reference to manuscript].") Since a large number of models and methods will be presented in the resulting manuscript, referring to the primary manuscript when discussing a specific model submitted to the competition would be uninformative. Second, a brief presentation of a specification with rationale is unlikely to make future publication of further work on the same methods difficult. Inherent in the design of this crowdsourcing endeavor, Drs. Craig and Rand-Hendriksen are coordinators of the surveys and submission process, not the authors of each specification. Third, each team leader will be required to agree at time registration to refer to the specific model submitted by another team in this competition if ever I incorporate their model in my future work. This is not enforceable, but strongly delineates norms regarding intellectual property set by this competition.

THE CONFIRMATORY STAGE

Fielding the Confirmatory Study: Once the working paper is submitted to the EQ working paper series (and potentially the EQ Plenary), we will field the confirmatory study, which is identical to the exploratory study except in that it includes the confirmatory pairs (see above).

Prior to the EuroQol Plenary, Appendix 1 of the working paper will be augmented to include one new column (Confirmatory predictions) and one new row (chi-square for each team's predictions). Also, the predictions (columns) will be re-arranged by the teams' chi-square. If a tie exists, the specification with the least number of parameters will be considered as the better analytic approach. Appendix 2 will also be rearranged by the teams' chi-square.

Knowing that the submissions will be ranked by their chi-square (1 criterion), each team is incentivized to provide their best set of predicted values (along with their model and estimation technique) based on this criterion. Although the selection of this measure of fit (or any other measure of fit) is arbitrary, the process will be fair, transparent, and well communicated in advance of the submissions and validation study. Other comparisons may be made between

the submissions and/or the validation evidence, but for discussion purposes only – the “winner” (i.e., 1st place) will be declared on the basis of the smallest chi-square.

At the EuroQol Plenary, the winner will be announced either via email or publically as part of the EQ Plenary (if the abstract is accepted). The notification will be sent/delivered by Drs. Craig and Rand-Henriksen (i.e., discussant). Also, the 2 augmented Appendices will be posted on the IAHPR website along side the link to the working paper, which will be on the EuroQol website.

While we do suggest that one model will be declared best in this particular competition, we are not suggesting that a specific model will be promoted as being universally best, or true in some deeper sense. The main objective is to get as many possible modeling approaches as possible on the table, so that their merits and weaknesses can be discussed. Furthermore, we are asking the teams to describe the process or rationale on which they have selected their model, and this, as we see it, may be of greater importance than the specifics of the models and estimation techniques; what criteria or methods do the different teams suggest are the best or most appropriate for selecting an analytical approach when they have been given a common dataset?

Nevertheless, the victorious team leader will be awarded (1) a small trophy to be designed by Drs. Craig and Rand-Hendriksen, (2) the option of first authorship of the manuscript along with Drs. Craig and Rand-Hendriksen, conditional on participation in the writing of the discussion and fulfillment of the Vancouver criteria, and (3) the privilege of writing the concluding paragraph of the manuscript. All of the other members of the winning team will be included in the acknowledgements. All other teams will be listed in the acknowledgements, and the names of the leader and members of each team will be listed in Appendix 2.

A concern that has been raised relates to inherent advantages of experience with US preferences, analyses and the survey instrument. To avoid potential conflicts of interest, Dr. Craig will post and distribute his submission prior to the receipt of any other submissions (placing his submission at disadvantage). This can serve as an example and allows him to review the submissions of others without concern that he might modify of his own. This review of submission forms is solely based on completeness and formatting, similar to his role as Co-editor of the EQ Working Paper Series. In addition, a runner up will be incorporated in the final manuscript if either Dr. Craig or Dr. Rand-Hendriksen is victorious. This guarantees that the final manuscript will have 3 authors.

On September 17 (after the EuroQol Plenary), this working paper will be minimally augmented, specifically: (1) The original Appendices will be replaced with the augmented Appendices. (2) Any relevant details concerning the fielding of the confirmatory study will be added to the results. (3) A scatter plot showing the exploratory and confirmatory chi-square results will be included as a way to delineate the fit of the alternative specifications. (4) As a form of sensitivity analysis, the results section will describe the extent to which the specifications predict the choices between mild (no 4s or 5s) vs. severe (no 1s or 2s) states and short (1 to 5 years) vs. long (5 to 10 years) durations. (5) The discussion section will incorporate any new changes and limitations. (6) Lastly, the team leader will submit his/her conclusion (i.e., the final word).

After these changes, the primary paper will be promptly submitted for peer-reviewed publication. Its publication is guaranteed, but the specific journal or issue is not yet known. Soon after initial submission of the primary manuscript, Drs. Craig and Rand-Hendriksen will submit their final report to the EuroQol Research Foundation. Aside from the sample probabilities from the confirmatory study in Appendix 1, the data from the confirmatory study will not be distributed until the final manuscript is published. All data and code will be made available for download on the IAHPR website shortly after its publication.