# Abstract

Background: Clinical parameters incorporated in health economic models are sometimes known to be monotonically related, meaning that for two parameters no estimate of one of the parameters should be greater than the estimate of the second parameter. Independent sampling of each parameters from known distributions could produce violations of the monotonicity assumption. A comparison of alternative methods for preventing or minimising violations of monotonicity has not, to our knowledge, previously been published.

Objective: To compare the appropriateness of ten chosen methods which could be used in when sampling from monotonically related parameters.

Methods: Summary statistics (mean and SE) were generated based on hypothetical individual patient data (IPD) for two parameters known to be monotonically related. The ten methods compared use only these summary statistics. The joint uncertainty about the mean values produced by each of the methods is compared with that derived directly from the IPD using a bootstrapping procedure. The similarity between the joint uncertainty based on the IPD, and those of the methods, is assessed. For those methods which produced appropriate joint estimates, a supplementary analysis is performed to their performance in situations where the summary estimates are based on different hypothetical populations.

Results: Threeof the ten methods were found to produce joint estimates of the parameters which did violate the monotonicity assumption, and did not introduce any noticeable discontinuities and biases into the joint parameter estimate. This included two types of covariance-based method, and a ‘difference method’ where one of the parameters is modelled using a two-stage process. In the supplementary analysisthe ‘difference method’ was shown to perform appropriately in this analysis, whereas one of the covariance-based methods produced violations of the monotonicity assumption.

Limitations A systematic review of methods used to take monotonicity into account was not undertaken. Only two hypothetical datasets were used to assess the results.

Conclusions The ‘difference method’ ensures that monotonicity cannot be violated, and also does not appear to introduce the discontinuities and biases observed when many of the other methods are used. It is also consistent in the results it produces and easy to use.

Key words:

Health economics

Decision analytic models

Parameter uncertainty

Monotonicity

# Introduction

## Background

The use of probabilistic sensitivity analysis (PSA) in cost-effectiveness models is strongly encouraged and widely used. (1,2) Relationships between parameters should be represented in the PSA. This paper discusses methods for handling one such type of relationship, namely parameters that are known or strongly suspected to be monotonically related, but not perfectly correlated.

An example of a monotonic relationship in health economics is when modelling a disease which has a less severe state and a more severe state. Assuming no comorbidities, and that both states were symptomatic, there are clinical reasons to believe that health related quality of life (HRQoL) in the more severe state will be lower than in the less severe state.

This paper compares ten different methods for estimating PSA parameters where the monotonicity assumption applies for two parameters. The inspiration for this paper was that we have observed authors of economic evaluations using naive and resampling approaches, which we show are inadequate for handling monotonicity.

## The Montonicity Assumption

The monotonicity assumption can be formally stated as follows: If there are M runs within the PSA, and the subscript i defines predicted values from the ith run, then U1i ≥ U2i for all i . If monotonicity were violated then some of the estimated values of U1 - U2 produced from the PSA would be negative.

# Methods

To test the methods hypothetical IPD were generated from 30 hypothetical patients, in which each individual’s HRQoL values were recorded while in a ‘worse’ health state U2 and a ‘better’ health state U1. We assume that means and standard errors of U1 and U2 would be reported, but the covariance would not be.

The ‘true’ joint uncertainty of the mean values of U1 and U2 was determined by bootstrapping: the pairs of observations from each of the 30 patients in the original dataset are repeatedly resampled with replacement, and the mean values of U1 and U2 recorded. Doing this 1,000 times produces a set of PSA results that were considered a gold standard against which the PSA produced by each of the methods can be compared. We assume that the more similar the PSA produced by one of the methods is to that produced by the bootstrapping procedure, the more appropriate it should be considered to be for handling this form of monotonicity.

Each method has access to the same basic summary statistics about the two parameters: the parameter sample means and standard errors. Although some methods involve assuming a correlation between variables, the true correlation between parameters is not used by any method, because it is not something that is commonly reported. The summary statistics reported are based on a dataset of hypothetical IPD. It is assumed that there are no missing observations, and so for each patient there is a record of their U1 state and U2 state.

## Simulated Data for main analysis

The individual patient data (IPD) are provided in the appendix in Table 1, and the corresponding scatter plot for these data is shown in Figure 1. Note that monotonicity applies in all cases as all points are below the diagonal line.

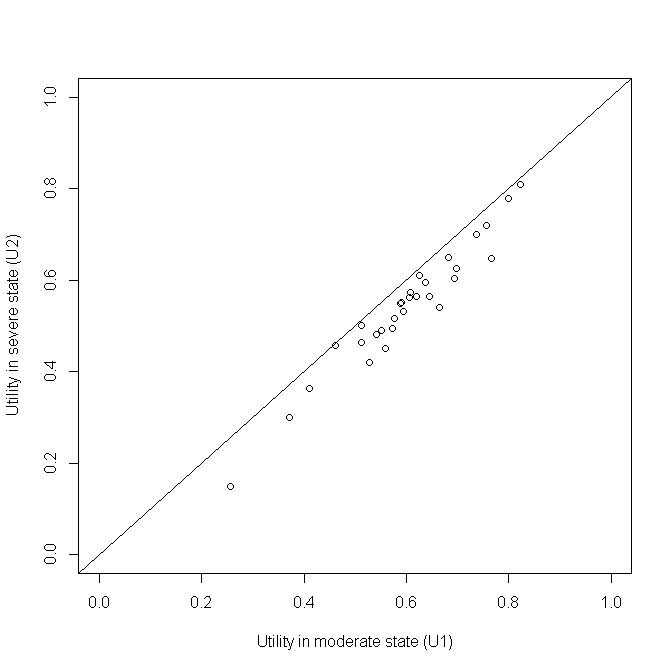


Figure 1 A plot of the simulated individual patient data

## Bootstrapped estimates of means

In cohort models the distribution used in PSA represent uncertainty around the mean value. PSA usually aims to represent uncertainty about the mean value of a parameter, rather than the full distribution of that parameter. Bootstrapped estimates produced the data shown in Figure 2. In this figure, no estimate of U1 is less than the corresponding estimate of U2. The two means are strongly but not perfectly correlated (r= 0.97). Because of this, there is some variability in the differences between the two estimates, U1 – U2, showing that simply adding E(U1) – E(U2), (i.e. 0.600 – 0.542 = 0.058) onto the PSA estimates of U2 to producing corresponding PSA estimates of U1 would not be correct.

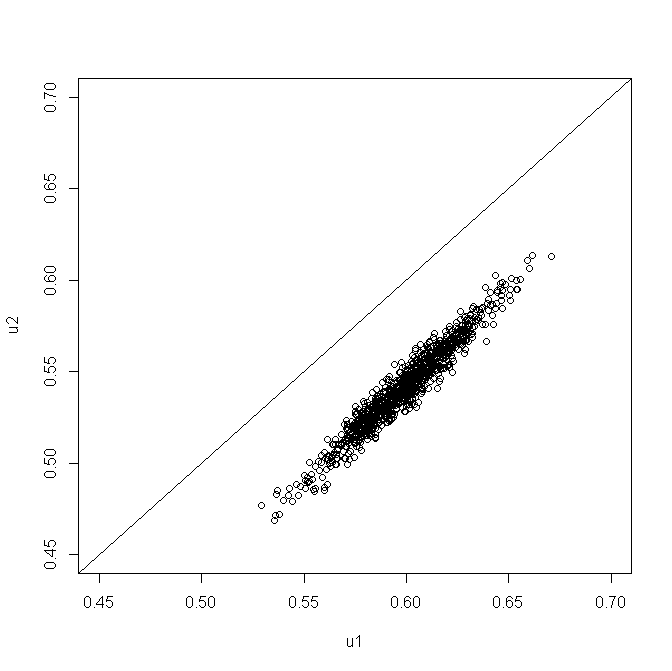


Figure 2 Scatterplot of 1000 PSA draws of the joint means of U1 and U2 produced by bootstrapping the IPD in table 1

## Summary statistics for main analysis

Table 2 shows the information assumed to be available to the modeller. In addition to this it is assumed that U1 is greater than U2.

|  |  |  |
| --- | --- | --- |
|  | **U1** | **U2** |
| Sample mean | 0.600 | 0.542 |
| 95% confidence interval of mean | 0.555 to 0.644 | 0.494 to 0.590 |

Table 1 The assumed available summary data. This is assumed to be the only information available to the modeller

## Summary statistics of supplementary analysis

A supplementary analysis was conducted for those methods shown in the primary analysis to produce appropriate results. In this supplementary analysis, it was assumed that estimates for the mean value of U1 and U2 were based on different populations. In our example, U1 is drawn from 80 patients and U2 is drawn from 15 different patients. For example, all 95 patients may be suffering from the same disease, but the 80 patients used to inform the U1 estimate are in a non-progressed state, and the 15 patients are in the progressed state. Neither the U1 values of the 15 patients when they were in the non-progressed state nor the U2 values of 80 patients when (or if) they will be in the progressed states are known. The summary data available are shown in Table 2.

|  |  |  |
| --- | --- | --- |
|  | **U1** | **U2** |
| Sample size | 80 | 15 |
| Sample mean | 0.600 | 0.542 |
| Standard deviation | 0.100 | 0.120 |
| Standard error | 0.011 | 0.038 |
| 95% confidence interval of mean | 0.578 to 0.622 | 0.481 to 0.603 |

Table 2 Assumed summary data available in supplementary analysis.

The supplementary analysis represents a situation often encountered by modellers, in which the smaller sample size of one or more of the distributions means that the confidence intervals of standard errors will overlap even though there are strong clinical reasons to believe the monotonicity assumption is correct.

## The Ten Methods

A systematic review of methods was not undertaken. Instead, the authors compiled a list of methods that they had personally undertaken, observed undertaken by modellers, or that appeared a potential solution. This produced a list of ten methods, which fit broadly into one of four classes:

1. **Simple and naïve methods** (methods one and two). A very naïve method is method one, independent sampling. A less naïve method is method two, in which the same random number approach is used;
2. **Resampling and replacement methods** (methods three to six), where draws from independent distributions are either selectively resampled or replaced with a value which ensures monotonicity;
3. **Multivariate model methods** (methods seven, eight and nine), where the variables used in the PSA are sampled jointly from a multivariate model where a covariance between variables is explicitly specified;
4. **Difference model methods** (method ten), where PSA draws for all but one of the variables are produced by adding a draw from a positively bounded distribution onto a draw for another distribution

The ten methods considered are described in Table 2. We use beta distributions for all methods except methods 7-9, as these methods use the bivariate normal distribution. The parameters of the beta distribution are derived from the summary information in Table 1. This is done by defining the sample mean as , the lower and upper 95% confidence intervals of the mean as and respectively, then estimating the sample variance ( as either or . The two parameters of the beta distribution are then calculated as and . Classes of methods where it is impossible, rather than just improbable, for monotonicity to be violated are marked with a † symbol. All methods were implemented using the R programming language. (3)

|  |  |  |  |
| --- | --- | --- | --- |
| **Class** | **Method Number** | **Name** | **Method Description**  For each of the PSA runs… |
| Naïve Methods | 1 | Independent Sampling | … take one draw from U1 and one draw from U2 independently (i.e. assume no covariance between U1 and U2) |
| 2 | Quantile Matching/  Number Seed Recycling | … use the same random number seed when drawing a sample from U2 and U1. (This is similar to selecting the same quantile from both distributions.) |
| Resampling and replacement methods† | 3 | Upward Replacement | Stage 1: draw a sample from U2  Stage 2: draw a sample from U1  Stage 3: Check if the value of U1 drawn is less than the corresponding value of U2 drawn. If it is, then replace the value of U1 with the U2 value. |
| 4 | Downward Replacement | Stage 1: draw a sample from U1  Stage 2: draw a sample from U2.  Stage 3: Check if the value of U2 drawn is greater than the corresponding value of U1 drawn. If it is, then replace the value of U2 with the U1 value. |
| 5 | Upward Resampling | Stage 1: draw first from U1.  Stage 2: draw from U2.  Stage 3: Check if the value of U1 is less than U2. If it is, then go back to Stage 2. If not, then stop. |
| 6 | Downward Resampling | Stage 1: draw first from U2.  Stage 2: draw from U1.  Stage 3: Check if the value of U2 is greater than U1. If it is, then go back to Stage 2 (i.e. resample). If not, then stop. |
| Multivariate model methods | 7 | AIVM Covariance | Assume that the covariance between U1 and U2 is equal to the average of the individual variances of the means (AIVM) of U1 and U2.  If assuming this covariance implies that the correlation between U1 and U2 is greater than 1, then instead select the covariance between U1 and U2 which implies a correlation of 1 between U1 and U2. |
| 8 | Lower Bounded Covariance Retrofitting | Select the minimum value of a covariance between U1 and U2 such that the two following conditions are met:  Condition 1: U1 – U2 > 0 for all PSA runs.  Condition 2: The covariance between U1 and U2 is greater than AIVM.  If this implies that the correlation between U1 and U2 is greater than 1, then instead use the covariance value associated with a correlation of 1. |
| 9 | Upper Bounded  Covariance Retrofitting | Methodology 8 but where the condition two is that the covariance between U1 and U2 is less than AIVM. |
| Difference model methods † | 10 | Beta Distribution Difference Modelling | We define U2=U1- Δ, where Δ~ Beta(a,b).  Stage 1: Calculate the beta parameter a and b  where and denote the mean and variance of the beta distribution, respectively with  **For variance of U1<variance of U2:**  Stage 2: Draw Δ from the beta distribution.  Stage 3: Draw U1 from the beta distribution.  Stage 4: Samples of U2 is calculated using samples of U1 minus samples of Δ.  **For variance of U1>variance of U2:**  Stage 2: Draw Δ from the beta distribution.  Stage 3: Draw U2 from the beta distribution.  Stage 4: Samples of U1 is calculated using samples of U2 plus samples of Δ. |

Table 3 Summary of the ten approaches considered. The † symbol indicates classes of methods where monotonicity cannot be violated.

## Simple and naïve methods

Method one, independent sampling, is the simplest method of all, and does not take the monotonicity condition into account. Nevertheless, in cases where the means of U1 and U2 are far apart and the standard errors of both parameters are small, this method may still produce PSA values which do not violate the monotonicity assumption. Method two involves using the same random number when drawing from both the U1 and U2 distributions.

## Resampling and replacement methods

Methods three to six are relatively simple to implement. All four methods involve sampling one of the two paired values, U1i or U2i, independently, before sampling the value, U2i or U1i. For methods three and four, the second value is then replaced with the first value if it violates the monotonicity assumption. For methods five and six, the second value is retained if it does not violate the monotonicity assumption, and resampled if it does violate the assumption. The second value is resampled until a value which does not violate the monotonicity assumption is drawn.

There are theoretical reasons to be concerned with both the resampling and replacement methods. The replacement methods can be shown to produce biased estimates of the mean value. Any systematic increase (or decrease) in the sample value will result in the average of 1,000 samples being greater than (or lower than) the true distribution mean. This phenomenon occurs independently of whether the value is set equal to the previously sampled parameter value, or whether it is resampled until monotonicity is upheld, although the bias will be less in the former methodology. Despite the known bias of these methods they have been included to formally highlight this.

## Multivariate model methods

Methods seven to nine involve selecting covariances on the basis either of the variances presented in the summary statistics for U1 and U2, or on whether monotonicity is maintained on all runs of the PSA. Method seven involves setting the covariance between U1 and U2 to the average of the individual variances of the means (AIVM). Method eight involves setting the covariance to such a value that no PSA draws violate the monotonicity assumption, subject to the constraint that the covariance is also greater than the AIVM. For method nine, the covariance is also set such that no PSA draws violate the monotonicity assumption, but this time subject to the constraint that the covariance is less than the AIVM.

A further logical constraint also applies to all three methods. This is that the covariances cannot imply a correlation with a magnitude greater than 1. The correlation of two random variables X and Y is defined as follows:

For this reason, the upper limit of the covariance must be . For approach seven, this effectively states that the covariance selected is:

This constraint also places an upper limit on the range of covariances which may be considered in methods eight and nine.

Both methods eight and nine involve selecting the most appropriate covariance by generating a ‘training set’ of 1,000 PSA draws from a bivariate distribution assuming the smallest acceptable covariance parameter. If any of the PSA draws violate monotonicity then the covariance parameter is increased slightly and another training set of 1,000 draws generated. The covariance is selected either if none of the draws in the training set violated monotonicity, or if the covariance could not be increased further without violating the other condition or the logical constraint described above.

Unlike the other approaches used here, these three approaches involve sampling from bivariate normal distributions rather than beta distributions, which is less appropriate in theory as the normal distribution, unlike the beta distribution, is not bounded to produce values between 0 and 1.

## Difference model method

The concept of this method is to find transformations of U1 and U2, such that the transformed variables are judged to be independent. We introduce a variable , which is the difference between U1 and U2. We assume that U1 is independent of and U2 is also independent of . Depending on the magnitude of the variance of U1 and U2, we define two different difference models. This is required to ensure we will always be able to calculate the variance of .

Since in our example both U1 and U2 are in the range between 0 and 1, the difference needs to be bounded between 0 and 1 as well. We assume a beta distribution Beta(a,b) for so that this condition is met. We calculate the two beta parameters a and b using the mean and variance of U1 and U2 as described earlier. In the PSA, we firstly draw from Beta(a,b), then draw either U1 or U2 from its normal distribution depending on the model used. Finally, we calculate samples of U2 using U2=U1- if the samples of and U1 have been drawn, or calculate samples of U1 using U1=U2+ if samples of and U2 have been drawn.

## Methods where monotonicity cannot be violated

For some of the methods, monotonicity cannot be violoated. These methods are three, four, five, six and ten. For methods seven, eight and nine, which use algorithms to select covariances between parameters, it is possible that for some runs monotonicity may be violated. Where violation of monotonicity is possible, modellers should be able to specify what level of monotonicity violation is tolerable. For example, monotonicity violation may be acceptable, so long as it occurs with a frequency of less than 1/10,000. For brevity, methods three, four, five, six, and ten will be described as satisfying ‘strict monotonicity’; whereas methods seven, eight and nine will be described as satisfying ‘relaxed monotonicity’.

## Comparing between methods

We use two visual approaches to compare the ten methods with the bootstrapped estimates based on the IPD. In all cases, we assume that the closer the output from a method is to the bootstrapped estimates, the better it is at accurately representing the relationship between U1 and U2 given only summary data.

Firstly, we produce scatterplots of 1,000 joint estimates of U1 and U2 for each of the ten methods. These are drawn on the same scale as the scatterplot shown in Figure 2, and so the joint patterns of scatter produced by each method can easily be compared with Figure 2.

Secondly, we use violin plots to compare the distribution of the quantities U1, U2, and U1 - U2 for each of the ten methods with the bootstrapped estimates. Violin plots are similar to box plots but also present kernel density estimates of distributions of the type presented in Figure 3. (4) An appropriate method for representing the monotonic relationship given only the summary data should produce distributions for these quantities which look similar to the bootstrapped values for U1, U2, and U1 - U2.

# Results

## Parameterisation of methods seven, eight, and nine

The average of the individual variances of the means (AIVM) is 0.000552, but the product of the two sample standard deviations is 0.000550. As the product of the two standard deviations defines the covariance at which the correlation is 1, and the correlation cannot be greater than 1, method seven becomes equivalent to assuming perfect correlation between U1 and U2. In method eight, in which the AIVM defines the lower bound of the range of covariance values which may be considered, a covariance of 0.000504 was identified, implying a correlation of 0.92. For method nine, in which the AIVM defines the upper bound of the range of covariances to search through (except where AIVM is greater than the product of the standard deviations), a covariance of 0.000360 was identified, implying a correlation of 0.65.

## Parameterisation of method ten

Given the summary statistics of U1 and U2 in Table 2, the beta parameters are a=33.02 and b=536.33. Figure 3 below shows the distribution of 1000 draws from U2 using U2=U1+ Δ alongside 1,000 draws of U1 and U2. We see that the distribution of U2 closely matches that of U2 from the data. The variations are due to sampling errors.

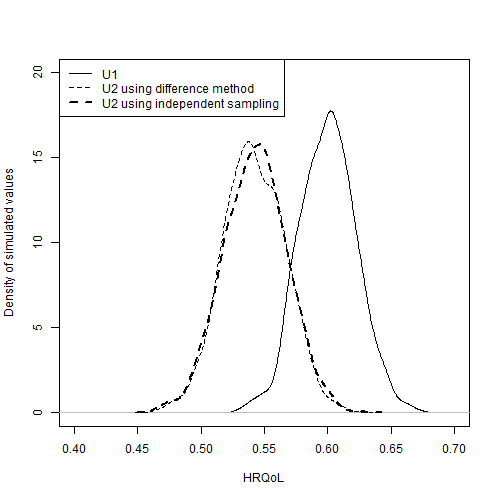


Figure 3 Comparison of the distribution of estimates of U2, U1, and U1\* produced using Method 10

## Scatterplots

In Figure 4 the scatterplots of U2 against U1 are shown for each of the 10 methods. The scatterplot from bootstrapping the IPD, shown in full size in Figure 2, is reproduced in Figure 4a for comparison. The diagonal line indicates parity between corresponding draws of U1 and U2. Scatter above this diagonal line shows that some proportion of the draws produced by the method violate the monotonicity assumption. A good method should be able to produce a similar pattern of scatter given the aggregate data as the bootstrapped method is able to produce using the IPD.

Figure 4b shows the scatterplot for method one. This shows some scatter above the diagonal line, highlighting the inadequacy of the approach. All other approaches appear to produce no estimates which violate the monotonicity assumption.

Methods three (Figure 4d), four (Figure 4e), five (Figure 4f) and six (Figure 4g) all show nonlinearities in the scatter, with no values above the diagonal line but relatively high densities of values just below the diagonal line. These discontinuities prove that these methods of ensuring monotonicity are liable to produce biases in the estimated mean values.

Most approaches produce patterns of variance in the scatter dissimilar to the bootstrapped scatter. Methods one (Figure 4b), three (Figure 4d), four (Figure 4e), five (Figure 4f), and six (Figure 4g) all produce uncorrelated scatter that is too wide, showing that the correlation is too low. By contrast methods two and seven produces scatter which is too narrow, as in this case the method is equivalent to assuming perfect correlation between U1 and U2.

We see from the scatter that methods eight (Figure 4i) and ten (Figure 4k) are closest in appearance to the bootstrapped scatter.

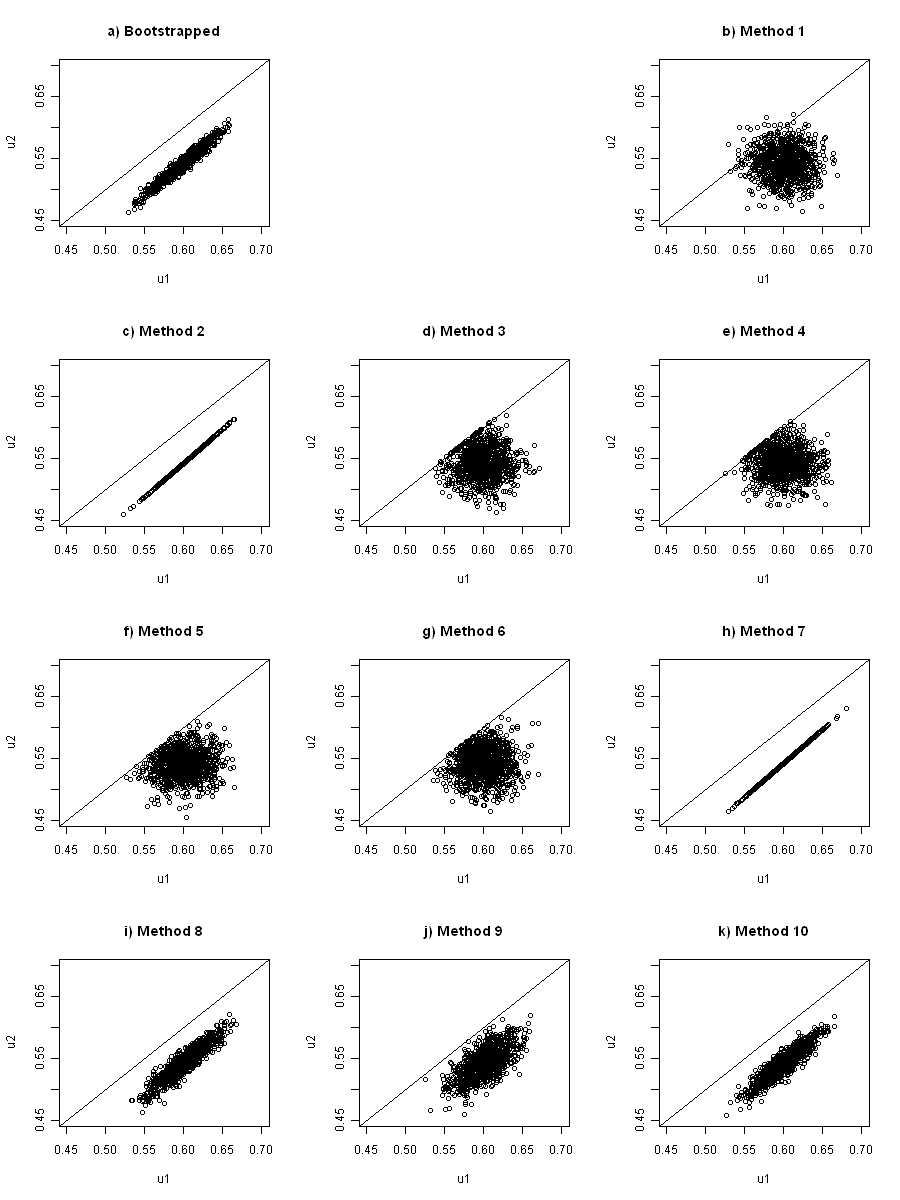


Figure 4 Scatterplots of U1 against U2 for each of the methods

## Monotonicity violation

The only approach where there was violation of monotonicity was method one (independent sampling). For this method 53 out of the 1000 PSA samples violated monotonicity. The precise proportion of samples violating monotonicity will differ slightly each time PSA is performed, due to stochastic uncertainty.

## Comparing U1, U2 and U1-U2

This section will compare the distribution of estimates of three quantities, U1, U2, and U1 – U2, produced by each of the ten methods, compared with the assumed gold standard, the bootstrapped data, allowing nuanced comparisons between the distributions to be made.

The top and middle subfigures of Figure 5 show the distributions of estimates of U1 and U2 respectively. They show that all methods appear broadly adequate in representing these quantities, in that all distributions have similar shapes. There is some indication showing that the resampling and replacement methods, methods three to six, produce biased means, in that the centres of the estimates, indicated by the white dots, do not line up with the bootstrapped centre, indicated by the horizontal dashed line. However in our example these differences are relatively small.

The bottom of the three subfigures, Figure 4c, shows the distribution of estimates of U1 – U2, i.e. the differences in paired draws of U1 and U2. Method one, independent sampling, producing some estimates where monotonicity is violated. Method two also produces some samples where monotonicity is violated, but it is also evident that most estimates produced by this distribution are within a small range, indicated by the very small length of the black line for this method compared with many other methods.

Two further problems are also observed. The first is a severe underestimation of the true uncertainty in this quantity, evident most strongly in methods seven and two. The second problem is evident in method three, four, five and six, which introduce a discontinuity into the distributions at the lower end (U2 – U1 = 0), while showing too wide a distribution at the upper end.

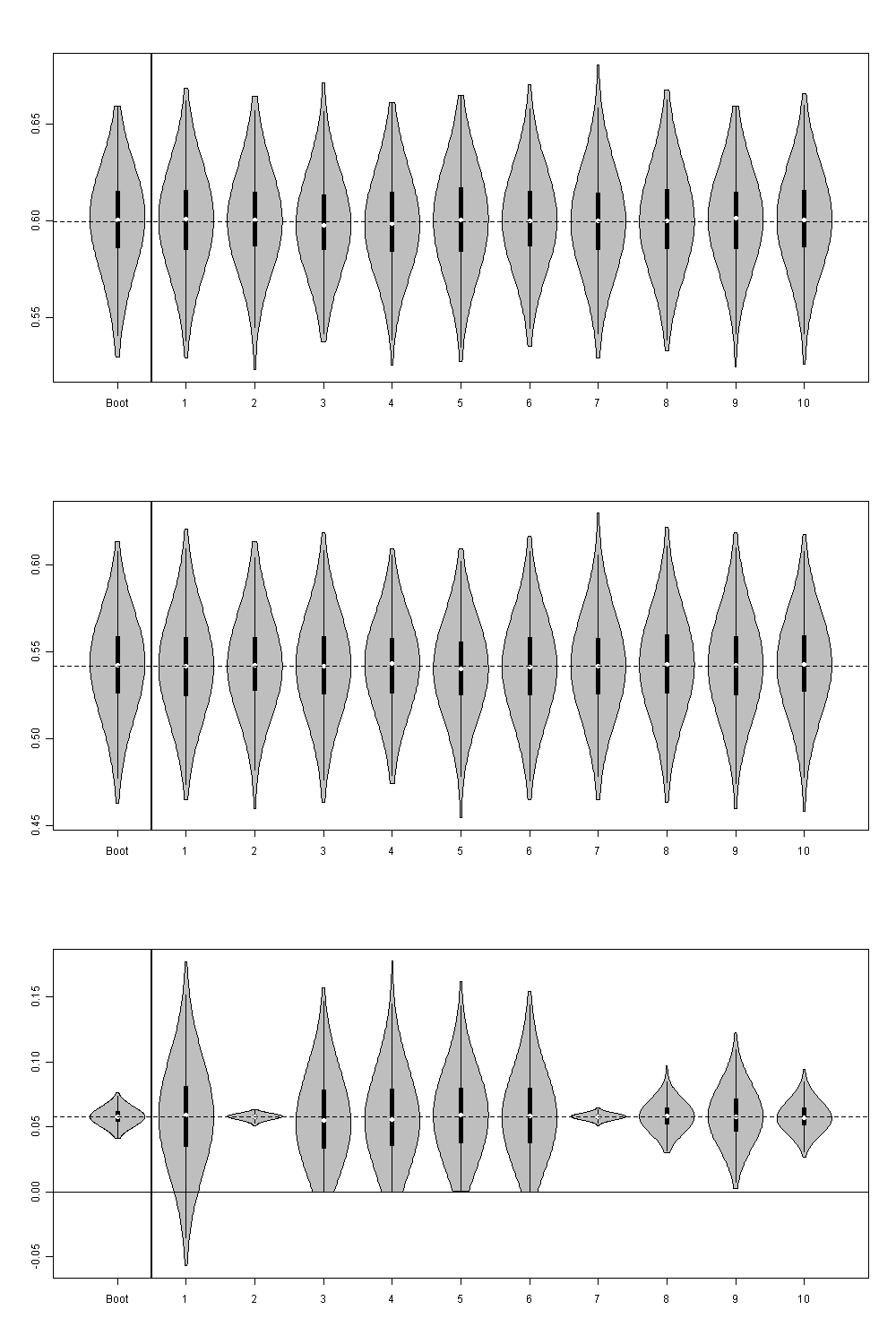


Figure 5 Violin plots of a) (top figure) distribution of estimates of U1; b) (centre figure) distribution of estimates of U2; c) (bottom figure) distribution of estimates of U1 - U2. In each plot distributions produced by each of the ten methods are compared with those produced by the bootstrapped estimates, labelled ‘Boot’. The white dots indicate the sample means. The thick vertical dark black lines indicate the interquartile range. The thin vertical black lines indicate the 95% intervals.

## Supplementary analysis

Based on these results, methods 8, 9 and 10 were evaluated further by seeing what outputs they produced when the alternative summary statistics were used. Here, U1 and U2 are derived from different samples of individuals, and so no IPD is available against which the results can be compared. The results of this additional analysis are shown in Figure 6. Figure 6a indicates that, unlike the results shown in Figure 3, the densities of the independently sampled estimates of U2 and the estimates of U2 created using the difference method, labelled U2\*, are no longer very similar. Instead the distribution of U2\* is wider than of U2.

Figure 6b shows the scatterplot produced by the first variation of the covariance methods, method 8. The distribution of scatter over this bivariate surface is narrow, and is unlike any of the subfigures in Figure 4 in that it does not run parallel to the diagonal line indicating parity. Instead, there is greater variation along the u2 axis than the u1 axis, as should be expected given that u2 has a greater standard error in this example. Figure 6c show the equivalent results for the other variation of the covariance method, method 9. This is similar to the scatterplot for method 8, but appears slightly narrower. It also has at least one point which is above the diagonal line, indicating that this method has violated the monotonicity assumption in this example.

The scatterplot for method 10 is shown in Figure 6d. The pattern of scatter appears slightly asymmetrical, and to vary more along the u1 dimension for smaller values of u2. No obvious discontinuities are introduced, and no violation of the monotonicity assumption can occur.

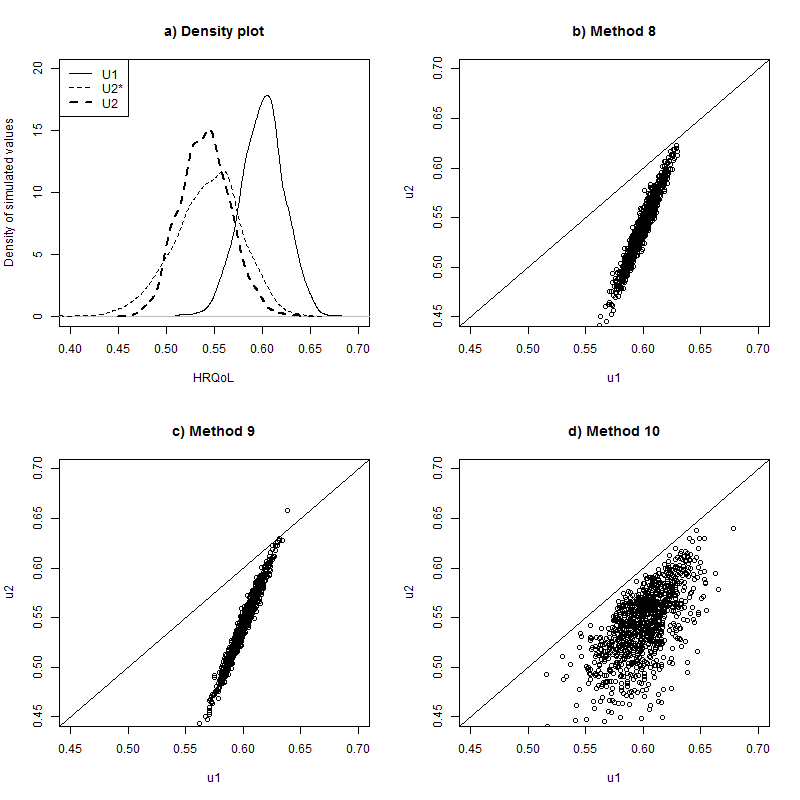


Figure 6 Results of supplementary analysis for selected method. a) Density plot of distribution of estimates for U1, U2 (independent sampling), and U2\* (estimates of U2 using the differencing method). b) scatterplot of u2 against u1 for method 8; c) scatterplot of u2 against u1 for method 9; d) scatterplot of u2 against u1 for method 10.

# Discussion

## Findings

This paper compared ten methods which may be used to handle the monotonicity assumption within PSA, against a ‘gold standard’ of bootstrapped estimates of hypothetical IPD. It confirmed that independent sampling is liable to produce violations of the monotonicity assumption, and so should not be adopted where it is important to incorporate this assumption within the PSA. It also found that a number of other commonly used methods for incorporating the monotonicity can effectively discard or misrepresent an important form of uncertainty: i.e. uncertainty about the difference between U1 and U2. Methods three, four, five, and six introduce implausible discontinuities into the distribution of differences between values; there are also theoretical reasons why these methods will produce biased means when monotonicity is violated by independent sampling.

Based on the results presented, and in particular the results shown in Figure 5c, only methods eight, nine and ten appeared to be broadly appropriate in their representation of both intra-distribution uncertainty (U1, U2) and inter-distribution uncertainty (U2 – U1).

Of the three methods that appeared appropriate, method ten has three clear advantages over methods eight and nine. Firstly, it uses statistical distributions (beta distributions) which are more appropriate for representing utility values than bivariate normal distributions used in methods eight and nine. Secondly, it is analytically impossible that method ten will produce any pairs of estimates which violate the monotonicity assumption, whereas occasional violation is possible with estimates produced by methods eight and nine. Thirdly, method ten is easier to implement, and can be implemented with far greater consistency, than methods eight and nine. Methods eight and nine both required relatively complex code to estimate the covariances in an automated way, and produce estimates which are affected both by simulation uncertainty and the size of the training sets of samples used to calibrate the covariance values; these problems are described in more detail in the limitations section. By contrast, method ten is simple enough that it can be run in a non-macro enabled Excel worksheet, which is available to download from the Health Economics and Decision Sciences (HEDS) website, and will produce identical estimates of the Beta parameters each time.

## Limitations

This section will describe limitations with the current analysis. This includes: not looking at results for a range of hypothetical datasets; not presenting a hypothetical example with three or more states; not investigating how dependent the results from the covariance-based methods are upon the size of the ‘training’ sets; and using distributions which bounded the range of utility values between the range 0 to 1. Each of these limitations will be discussed in more detail.

The first limitation is that we did not look at results for a range of different hypothetical datasets with different individual level and summary characteristics. For example, in our primary analysis using hypothetical IPD the standard error of U1 and the standard error of U2 are similar, and this factor may have affected the results comparing each of the methods.

A second limitation, related to the first limitation, is that our hypothetical dataset have only two disease severity states, U1 and U2, rather than three states such as U0, U1, and U2, where the utility of U0 is expected to be greater than for U1, and U1 to be greater than for U2. Introducing further states would lead to complications for methods seven, eight and nine, for example, as we would have to make decisions about the covariance between parameters U0 and U2, as well as between U0 and U1, and between U1 and U2. This provides some further justification for choosing method ten.

A third limitation relates to how methods eight and nine have been implemented. Both of these involve choosing covariance parameters conditional of whether any pair of values in a ‘training’ sample of 1,000 draws violates the monotonicity assumption. As the size of the ‘training’ sample increases, the probability of extreme values, including values which violate monotonicity, increases, and so we should expect the covariance selected to depend partly on the size of the training sample used.

The final limitation is relatively simple to address. Because worse-than-death health states exist and are evaluated in some economic evaluations, it may be inappropriate to use estimates directly from a Beta distribution which is bounded within the range 0 to 1. This problem could be addressed by rescaling the output from the Beta distributions from the range 0 to 1 to the range -0.594 to 1, for example, if representing HRQoL scores using the EuroQol Quality of Life (EQ-5D) scale. (5,6)

## Implications for Research

Further research should look at the dependence of the results and conclusions on the data we have used. For example, in our hypothetical dataset no individual happened to report a higher HrQoL score in the worse state than the better state. Research could also consider the choice of statistical distribution on the results produced. Further research should also consider the generalisation of these methods to three or more state. For the difference method, this can be done by using the method iteratively. Additionally, further empirical research could be conducted to investigate the influence of the choice of method on decision uncertainty within the technology appraisal process.

## Implications for practice

The decision to use a method which forces monotonicity needs to be based on good clinical evidence and in consultation with clinical experts. If there is neither strong evidence nor a consensus of clinical opinion that variables have this monotonic relationship, then simpler methods such as independent sampling, in which monotonicity can be violated, may be more appropriate.

Some of the methods for ensuring monotonicity should be avoided for statistical reasons. These are the replacement methods (three and four) and the resampling methods (five and six). These produce biased estimates of the means and variances of the distributions, shifting the centres of these distributions by introducing discontinuities. They produce estimates which do not adequately represent the summary information used in their construction. We warn against the use of these methods even where there is strong belief in the clinical relationship between variables.

If the decision to use a method which ensures monotonicity between variables is made, then the most appropriate methods appear to be either to be a covariance-based method or the difference method. As described in the findings section above, of these two approaches, the difference method has a number of advantages, and we have produced an easy-to-use Excel workbook to facilitate its use.

## Conclusion

We compared ten methods for producing PSA input parameters for two monotonically linked variables with a ‘gold standard’ which uses IPD. We identified a method which we deem to have superior properties in terms of representing uncertainty about the difference between variables, without producing biases or discontinuities in the simulated values. For this reason, we recommend the method be adopted within health technology appraisals where monotonicity between parameters is believed to exist.

# References

1. NICE. Guide to the methods of technology appraisal [Internet]. NICE methods guide. 2008 [cited 2012 Oct 15]. p. 80. Available from: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf

2. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health economics [Internet]. 2005 Apr [cited 2012 Jul 15];14(4):339–47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15736142

3. R Development Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2011. Available from: http://www.r-project.org/

4. Hintze JL, Nelson RD. Violin Plots: A Box Plot-Density Trace Synergism. The American Statistican. 1998;52(2):181–4.

5. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. PharmacoEconomics [Internet]. 2000 Jan [cited 2012 Dec 4];17(1):13–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10747763

6. Mann R, Gilbody S, Richards D. Putting the “Q” in depression QALYs: a comparison of utility measurement using EQ-5D and SF-6D health related quality of life measures. Social psychiatry and psychiatric epidemiology [Internet]. 2009 Jul [cited 2012 Dec 4];44(7):569–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19011721