Structure

Introduction

Background

In cost-effectiveness models, it is highly recommended that probabilistic sensitivity analysis (PSA) is performed. (Claxton et al., 2005; NICE, 2008) This requires producing numerous samples of parameter values used in the model, whose variability is intended to reflect uncertainty in the true value of those parameters. In many cases, the values of two or more parameters are known to be related to each other in some way, and this relationship also needs to be reflected appropriately in the PSA.

One type of relationship that could exist is that two or more variables are monotonically related. By this we mean that when considering two variables X and Y, though we are uncertain about both the true value of X and the true value of Y, we *are* certain that Y is greater than X. A common example of this is where a disease has a less severe state, and a more severe state, and it would be clinically implausible to assume that the mean health-related quality of life (HRQoL) while in the less severe state is lower than in the more severe state.

In this paper, we compare ten different methods for jointly simulating the PSA of two variables that we assume to be monotonically related. These ten methods fit broadly into one of four classes of method:

1) **Naïve methods** (methods one and two), where the two variables are sampled independently, or the same random number stream is used for simulating both variables;
2) **Resampling and replacement methods** (methods three to six), where draws from independent distributions are either selectively resampled or replaced with other draws;
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

In this paper we compare the properties of PSA samples created by each of the ten methods. All methods use only summary statistics, sample means and standard errors, which are often the only data available to modellers. In our comparison, the summary means and standard errors are derived from hypothetical individual patient data reporting the HRQoL for thirty patients with a hypothetical disease which could either be in a moderate state or a severe state. For each patient, we have produced values for their HRQoL in the moderate state, and also their HRQoL in the severe state. From the individual patient data (IPD) we produce 1,000 joint estimates of the mean HRQoL in the moderate state and in the severe state using a bootstrapping procedure. These bootstrapped estimates, based directly on the IPD, are the gold standard against which the estimates produced by each of the methods, which use only summary data, are compared. In general, we consider methods

which produce PSA samples most similar to the bootstrapped estimates to be preferable to those which produce dissimilar PSA samples.

The inspiration for this paper was that we have observed authors of economic evaluations using naive and resampling approaches, which we believe are inadequate for handing monotonicity in this context.

Method

Simulated Data

Our data is of thirty hypothetical patients who progress from a moderate disease state (Stage 1) to a more severe disease state (Stage 2). Each individual's HRQoL in the less severe disease state (U1) and the mere severe disease state (U2) is reported. The individual patient data (IPD) are shown in the appendix in **Error! Reference source not found.**, and the corresponding scatter plot for these data are shown in Figure 1.

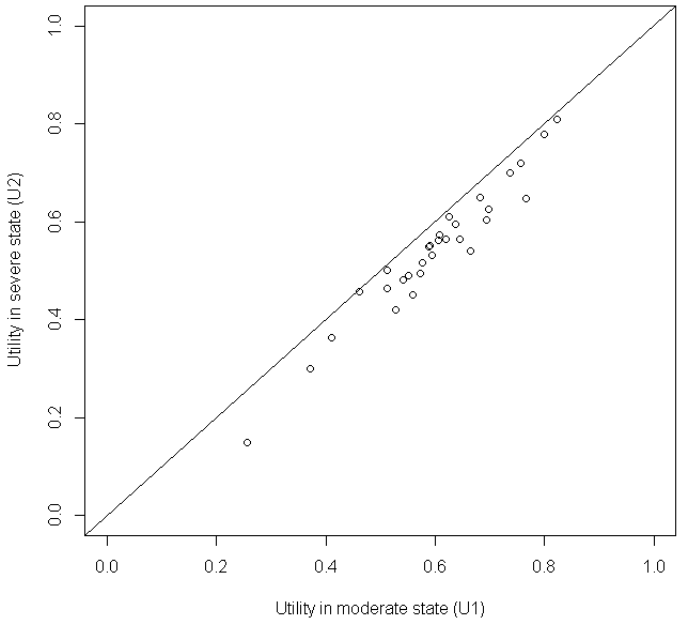

Figure 1 A plot of the simulated individual patient data

Bootstrapped estimates of means

As modellers are typically interested in representing uncertainty in expected values (uncertainty in the means) rather than predicted values (uncertainty and variability in the range of values encountered), 'true' uncertainty in the mean values of U1 and U2 was estimated by repeatedly resampling the IPD, and for each resample calculating the mean values of U1 and U2 produced. Doing this 1,000 times produced the data shown in Figure 2. This approach illustrates what the

modellers would be able to produce for the PSA if they had access to the IPD, and so represents the 'gold standard' against which the other methods, which have access only to aggregate level data, are compared.

We can see that the two parameters are monotonically related, as no estimate of U1 is less than the corresponding estimate of U2, and so no value crosses the diagonal line. We can also see that though 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. This shows 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, as it would not accurately represent the uncertainty in the differences between U1 and U2.

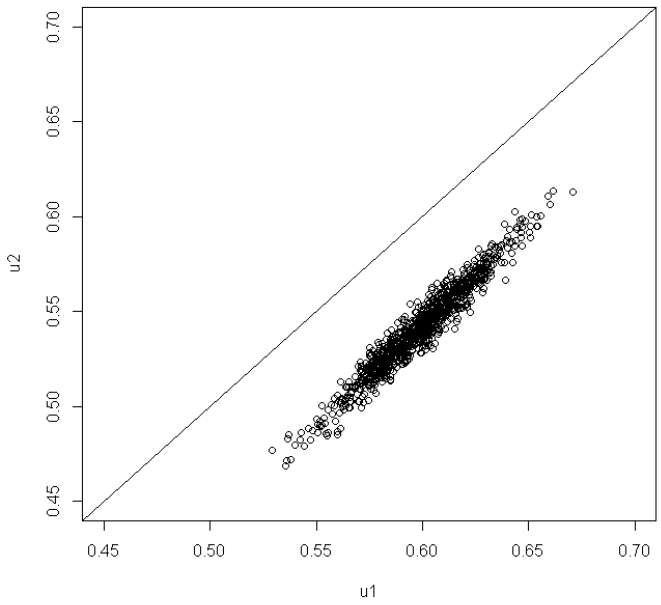

Figure 2 Scatterplot of 1000 PSA draws of the joint means of U1 and U2 produced by bootstrapping the IPD in table 1

Summary statistics

In our hypothetical example, we assume the modeller will not have access to the IPD, but only to the summary information shown in Table 1. This summary information, together with the knowledge that U2 should be less than U1, is the only information used in each of the ten approaches described below.

	U1	**U2**
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

The Montonicity Assumption

When modellers are generating multiple estimates for use within PSA using these summary data, the key monotonicity condition that must hold is that an estimate of U2 should always be equal to or lower than a corresponding estimate for U1. More formally, if there are M runs within the PSA, and the subscript i defines predicted values from the ith run, then $U1_i \geq U2_i$ for all i ; where M is the total number of PSA samples. If monotonicity were violated then some of the estimated values of U1 - U2 produced from the PSA would be negative.

The Ten Methods

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. 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. (R Development Core Team, 2011) The R code, including annotations, is presented in the appendix below.

Comment [JWM1]: How much detail should I go into describing this?

a = ((1 – mu) / var – 1 / mu) * mu ^ 2
b = a * (1 / mu – 1)

mu = sample mean
var = ((sample upper 95% interval – sample mean) / 1.96) ^ 2

Options:
1)Assume the reader knows all this;
2)Just reference something for the a, b identifies (If so, what?)
3) Provide the above definitions here
4) Provide the above definitions in an appendix

Which option should I go for?

Class	Method Number	Name	Method Description
Naïve Methods	1	Independent Sampling	For each of the PSA runs, 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	For each of the PSA runs, use the same random number seed when drawing a sample from U2 and U1. (This is equivalent to selecting the same quantile from both distributions.)
Resampling and replacement methods†	3	Upward Replacement	For each of the PSA runs: 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	For each of the PSA runs: 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	For each of the PSA runs: 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 (i.e. resample). If not, then stop.
	6	Downward Resampling	For each of the PSA runs: 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 second condition is that the covariance between U1 and U2 is less than AIVM.
Difference model methods †	10	Beta Distribution Difference Modelling	**For variance of U1<variance of U2:** 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 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:** We define U1=U2+ 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

Comment [JWM2]: Now using the beta distribution for the components.

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

Naïve methods

Methods one and two are both simple. Method one, independent sampling, is the simplest method of all, and does not take the monotonicity condition into account at all. 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. With the data considered here, however, this is not the case, and so the approach is liable to produce erroneous samples. Method two has been observed in economic evaluations, and involves using the same random number when drawing from both the U1 and U2 distributions. Method two is broadly equivalent to pairing the quantiles from the estimated distributions of U1 and U2 within PSA runs, matching the lowest estimate of U1 with the lowest estimate of U2, the second lowest estimate of U1 with the second lowest estimate of U2, and so on. For this reason, quantile-pairing was not considered as a separate strategy.

Resampling and replacement methods

Methods three, four, five and six have also been observed in economic models, as they are relatively simple to implement. All four methods involve sampling one of the two paired values, $U1_i$ or $U2_i$, independently, before sampling the value, $U2_i$ or $U1_i$. For methods 3 and 4, 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 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 methodologies. Despite the known bias these methods have been included to prove this for the novice reader.

Comment [JWM3]: Is there something I could reference for this?

Multivariate model methods

Methods seven, eight and nine each 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 also less than the AIVM. 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.

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

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

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

$$\operatorname{cov}(U_1, U_2) \leq \sigma_{U_1} \sigma_{U_2}$$

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

The R code used to implement methods seven, eight and nine is presented in the appendix.

Difference model methods

The concept of this method is to find transformations of U1 and U2, such that the transformed variables are judged to be in dependent. We introduce a new uncertain variable , which is the difference between U1 and U2. We believe 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 to make sure we will always be able to calculate the variance of .

Since 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. In 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.

For variance of U1<variance of U2

If variance of U2 is greater than the variance of U1, then let U2 = U1- Δ. Let the mean of Δ be and variance of Δ be , then

$$\operatorname{cov}(U_1, U_2) = \operatorname{cov}(U_1, U_1 - \Delta) = \operatorname{cov}(U_1, U_1) - \operatorname{cov}(U_1, \Delta)$$

The covariance between U1 and U2 is

.

For variance of U1>variance of U2

If variance of U1> variance of U2, then let U1=U2+ Δ. Let the mean of Δ be and variance of Δ be , then

The covariance between U1 and U2 is

.

From equation (1) and (2) if model U2=U1- Δ is used, or from equation (3) and (4) if model U1=U2+ Δ is used, the beta parameter a and b can be expressed as

The R code used to implement this method is presented in the appendix.

Methods where monotonicity cannot be violated

For some of the methods, it is analytically impossible for monotonicity to be violated, and so they must satisfy the monotonicity condition. 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 each other, and with the bootstrapped estimates based on the IPD. In all cases, 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. This comparison is facilitated by using violin plots, which are similar to box plots but also present kernel density estimates of distributions of the type presented in Figure 3. (Hintze & Nelson, 1998) 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 each of these three quantities.

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 parameter a=33.0224 and b=536.330. Figure 3 below shows the distribution of 1000 draws from U2 using U2=U1+ Δ alongside 1000 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.

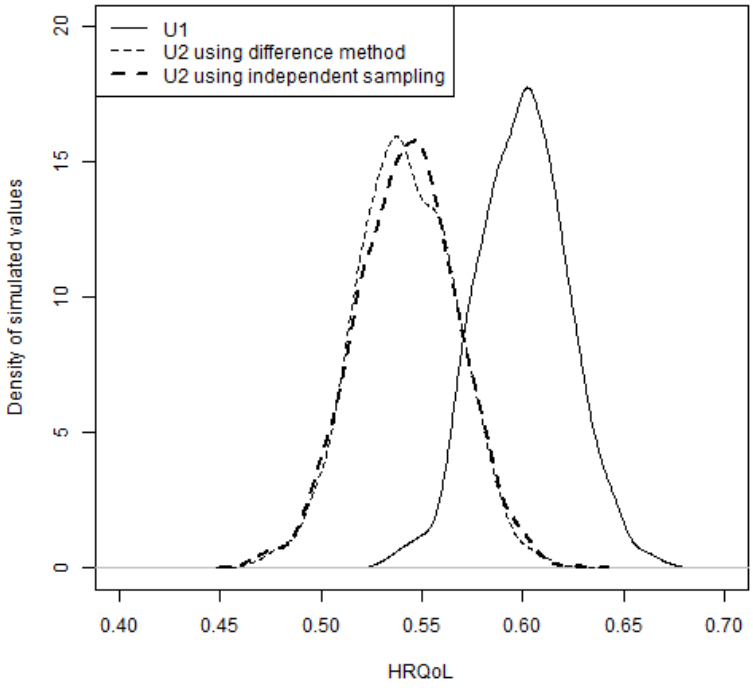

Figure 3 Comparison of the distribution of estimates of U2, U1, and U1* produced using Method 10

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.

Error! Reference source not found.b shows the scatterplot for method one. This shows some scatter above the diagonal line, showing that some proportion of the draws violates the monotonicity assumption, highlighting the inadequacy of the approach. **Error! Reference source not found.**c shows the scatterplot for method two, in which the majority of the scatter appears to follow a diagonal line parallel to the parity line, but a minority of the scatter does not, including some estimates where monotonicity is violated. All other approaches appear to produce no estimates which violate the monotonicity assumption.

Comment [JWM4]: Is/how much further discussion is needed?

Methods three (**Error! Reference source not found.**d), four (**Error! Reference source not found.**e), five (**Error! Reference source not found.**f) and six (**Error! Reference source not found.**g) 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 suggest that these methods of ensuring monotonicity are liable to produce biases in the estimated mean values.

The majority of the approaches appear to produce patterns of variance in the scatter which are qualitatively dissimilar to the bootstrapped scatter. Methods one (**Error! Reference source not found.**b), three (**Error! Reference source not found.**d), four (**Error! Reference source not found.**e), five (**Error! Reference source not found.**f), and six (**Error! Reference source not found.**g) all produce uncorrelated scatter that is too wide, indicating the correlation of the PSA estimates is too low. By contrast method 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 (**Error! Reference source not found.**i), nine (**Error! Reference source not found.**j) and ten (**Error! Reference source not found.**k) are closest in appearance to the bootstrapped scatter.

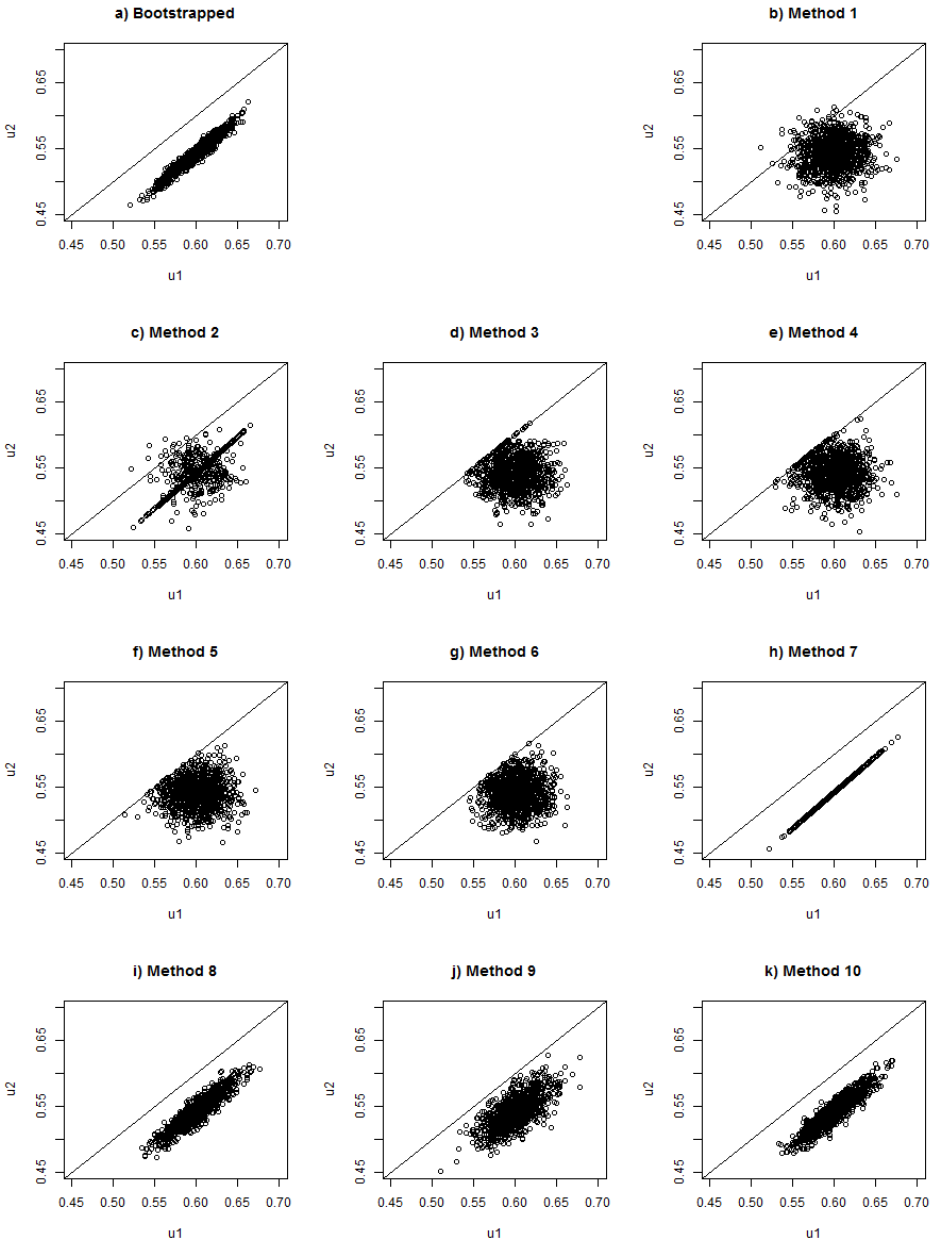

Figure 4 Scatterplots of U1 against U2 for each of the methods

Monotonicity violation

The two replacement methods (three and four), the two resampling methods (five and six), and the Beta method (method ten) are all designed such that it is analytically impossible for them to violate the monotonicity assumption. The other methods could all potentially violate monotonicity. In this example, the only approach where monotonicity is violated in the 1000 PSA samples being compared is method one, independent sampling, where 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 gold standard, the bootstrapped data. Like the scatterplots shown in Figure 2 and **Error! Reference source not found.**, they therefore allow 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 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, **Error! Reference source not found.**c, shows the distribution of estimates of U1 – U2, i.e. the differences in paired draws of U1 and U2. As shown in Figure 4, we see clearly that method one, independent sampling, producing some estimates where monotonicity is violated, because some of the distribution of values is below the 0. In our example method two is also shown to produce some samples where monotonicity is violated, although it is also evident that the majority of the estimates produced by this distribution are within a small range, as indicated by the very small length of the black line for this method compared with many other methods.

Two other types of problem are also observed. The first type of problem is a severe underestimation of the true uncertainty in this quantity, which is evident most strongly in method seven. The second type of 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.

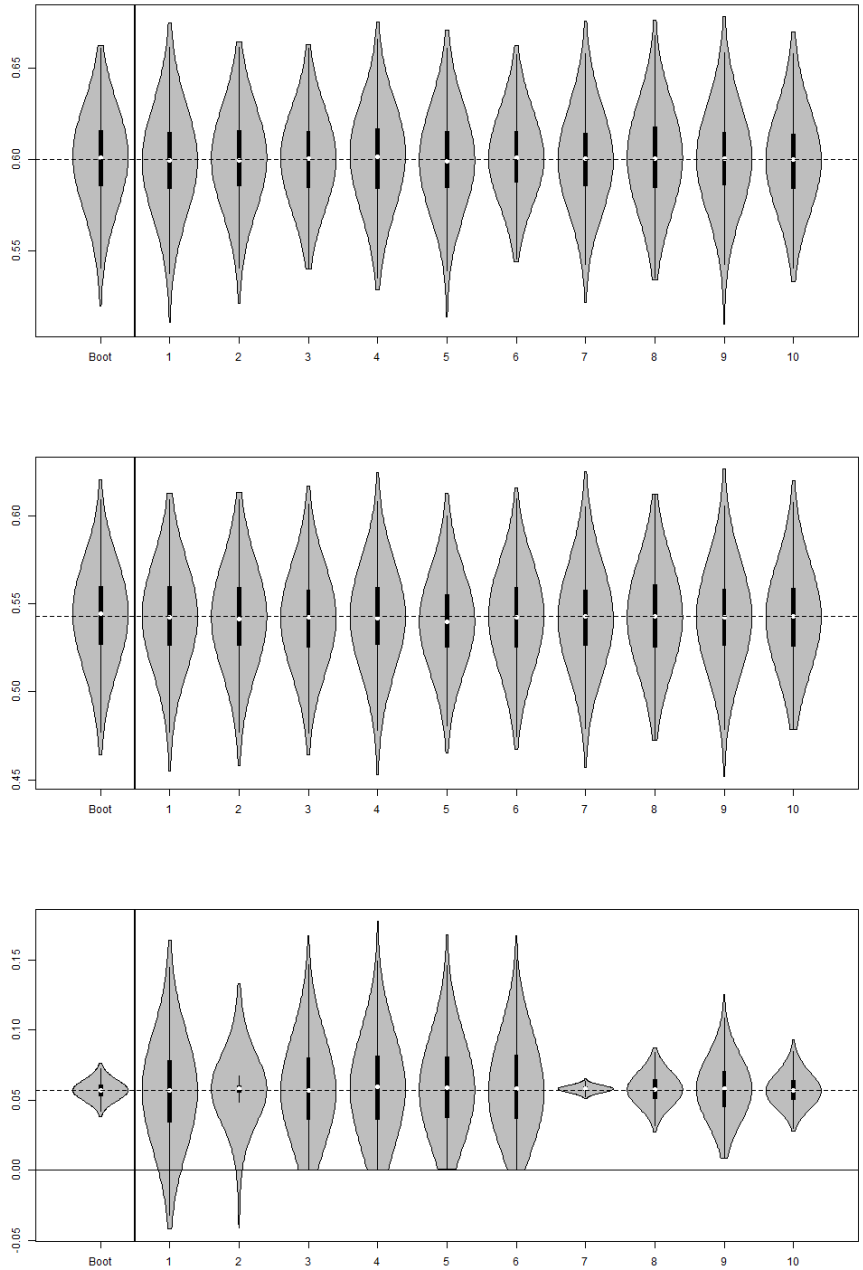

Figure 5 Violin plots of a) (top figure) distribution of estimates of U1; b) (centre figure) distribution of estimates of U2; c) (bottom figure) distribution of estimates of U1 - U2. In each plot distributions produced by each of the ten methods are compared with those produced by the bootstrapped estimates, 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.

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 estimates. 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 to assume that these methods will produce biased means, although such biases were not overly apparent in our results.

Based on the results presented, and in particular the results shown in Figure 5c, only methods eight, nine and ten appeared to broadly appropriate 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 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 below. By contrast, method ten is simple enough that it can be run in a non-macro enabled Excel worksheet [Ref], and will produce identical estimates each time.

Limitations

This section will describe some 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 looking at how dependent the results from the covariance-based methods are the size of the 'training' sets; and using distributions which bounded the range of HRQoL values between the range 0 to 1. Each of these limitations will now 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 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 second 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

HRQoL 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.

A third limitation we have identified 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 easily addressed by rescaling the output from the Beta distributions from the range 0 to 1 to the range -0.594 to 1, in the case of EQ-5D. [References needed.]

Implications for Research

Further research should look at the dependence of the results and conclusions on the data we have used.

Further research should look at the effect of representing lognormal or gamma distributions, rather than just normal and beta distributions.

Further research to look at how these methods can be generalised to three or more states.

Further research to see whether using method 10 in a previous HTA would have affected decision uncertainty.

[Need to think about this section again as we've now done both of the things we flagged for further research!]

Implications for practice

Although our analyses were performed in the statistical programming language R, method ten, which has better theoretical properties than the other methods and compares very well against the other methods in our example, can be easily implemented in Excel. Because of this, we recommend this method be used in practice.

In cases where the means for U1 and U2 are very close together and the standard errors are large, it is important to ask how confident we are about the validity of the monotonicity assumption, and how implausible it would be that the mean HRQoL in U1 is actually lower than the mean HRQoL in U2. If we are confident that the monotonicity assumption is correct, then method ten should be able to ensure that this relationship is represented in the PSA.

Conclusion

Within this paper, we have compared ten methods for producing PSA for two monotonically linked variables with each other, and with a 'gold standard' which uses IPD. We found a new method which

we developed 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 widely within health technology assessments.