# Abstract

Background: Assuming monotonically related variables are independent in health economic models can produce spurious parameter estimates and possibly suboptimal decisions. A comparison of methods for preventing or minimising violations of monotonicity has not, to our knowledge, previously been published.

Objective: To compare ten methods which could be used when sampling from monotonically related parameters.

Methods: Summary statistics (mean and standard error) 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, and the estimates produced compared with bootstrapped estimates which used the IPD. For those methods which produced appropriate joint estimates, further analyses using a separate dataset were performed.

Results: Threeof the methods were found to produce joint estimates of the parameters which did not 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, each involving only two monotonic levels, were used to assess the results.

Conclusions The ‘difference method’ ensures that monotonicity cannot be violated. It does not appear to introduce discontinuities or biases. It produces consistent results, and it is easy to implement.

Key words:

Probabilistic Sensitivity Analysis

Health Economics

Decision Analytic Models

Parameter Uncertainty

Stochastic Uncertainty

# 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 to be monotonically related.

As an example of a monotonic relationship in health economic evaluation, consider a hypothetical disease with two levels of severity: moderate and severe. Although there may be uncertainty about the health-related quality of life (HRQoL) associated with being in either state, there is also a strong expectation that a patient will have a lower HRQoL when the disease is in the severe state than the moderate state. If many estimates are produced independently for each health state, then on occasion the monotonicity assumption will be violated, meaning that some estimates of the HRQoL in the severe state are higher than corresponding estimates of the HRQoL in the moderate state. Despite this monotonic relationship, the two variables are not simple, analytic derivatives of each other with perfect correlation. Because of this, there is also need to appropriately represent stochastic variation and uncertainty in estimating both parameters.

This paper uses summary estimates derived from a hypothetical individual patient dataset (IPD) to compare independent sampling with nine other methods detailed in the methods section, which may be more appropriate for handling the monotonicity assumption. Results from each method were compared with the bootstrapped estimates from the IPD. Methods which performed better were assessed tested further using a second hypothetical dataset.

## The Montonicity Assumption

The monotonicity assumption can be stated more formally as follows. We define as the HRQoL in health state 1, and as the HRQoL in health state 2. Our assumption is that health state 1 is associated with higher HRQoL than health state 2, and so our expectation is that ≥ . Due to estimation uncertainty, we do not know the true value of either or , but instead have a range of estimates for and for , such that , where , refers to a stochastic function, and to parameters used within that function. Within PSA, M realisations of and M realisations of will be generated. If indexes the ith realisation of , then the monotonicity condition requires that , or equivalently , for all *i* .

# 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 and a ‘better’ health state . We assume that means and standard errors of and would be reported, but the covariance would not be.

The ‘true’ joint uncertainty of the mean values of and 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 and recorded. Doing this 1,000 times produces a set of IPD-derived PSA results. These IPD-derived PSA results were considered considered the 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 likely it should be considered appropriate to be for handling this form of monotonicity. 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.

The IPD are provided in the appendix A in Table A1, 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 about here]

## 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 on the difference in expectations to U2 to estimate U1 would be incorrect.

[Figure 2 about here]

## Summary statistics for main analysis

Table 1 shows the information assumed to be available to the modeller. In addition to this it is assumed that *U*1 is greater than *U*2.

[Table 1 about here]

## 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.

[Table 2 about here]

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 Methods

A systematic review of methods was not undertaken. Instead, we compiled a list of methods that we have used ourselves, seen others use, or that appeared a promising potential solution. This produced a list of ten methods including independent sampling, which fit broadly into one of four classes: Naïve/simple methods; resampling and replacement methods; multivariate model methods; and difference model methods. Each of these four classes of method will be described in more detail below.

The ten methods considered are described in Table 3. Classes of methods where it is impossible, for monotonicity to be violated are marked with a † symbol. All methods were implemented using the R programming language. (3) The R code used to perform the analyses is included as appendix B. The Difference Model method has also been implemented in Excel, and is included as supplementary material.

It may be useful to consider the ten methods with regard to the distinction made previously between the functions used to generate the estimates ( and the parameters passed to those functions (). For example, the resampling and replacement methods (methods three to six) do not introduce any dependence between and , but allow instead allow to be drawn from either or (where k’ = 1 if k=2 and vice-versa). Conversely, the multivariate methods (methods seven to nine) link and through a covariance structure, but the choice of covariance parameter is dependent on sampled estimates of and , including the number of samples used, M. Method ten, like methods seven to nine, also links and , but does not use a covariance structure to do so.

**Method one, independent sampling**,

This 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 Quantile matching**

involves using the same random number when drawing from both the U1 and U2 distributions which ensures monotonicity when the overlap between the distributions is confined to the upper tail of one distribution and the lower tail of the second distribution.

Method three Upward replacement and Method four downward replacement

They involve sampling one of the two paired values, U1i or U2i, independently, before sampling the corresponding values U2i or U1i. For methods three and four, the second value is then replaced with the first value if it violates the monotonicity assumption.

Method 5 upward resampling and method 6 downward resampling

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 in this paper 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 if either 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. Because of the use of the training set, the covariance parameter is in theory dependent on the sample size M, because the probability of observing violation of monotonicity () at least once increases with sample size. The implications of this are discussed further in the discussion section.

Unlike the other approaches used here, methods, seven, eight and nine, 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 both U1 U­2 are 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 so that this condition is met. We calculate 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 1, and so the joint patterns of scatter produced by each method can easily be compared with Figure 1.

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

This section is divided into three subsections. The first subsection details the parameterisation of methods seven to ten. The second subsection compares the ten methods using the first hypothetical dataset, in which U1 and U2 are drawn from the same patient population. The third subsection compares the three most promising methods using the second hypothetical dataset, in which U1 and U2 are assumed to be derived from different patient populations.

## Parameterisation

### Parameterisation of methods seven, eight, and nine

The average of the individual variances of the means (AIVM) is 0.000552, but as the product of the two sample standard deviations is 0.000550 the value is set to 0.000550, equivalent to assuming perfect correlation between U1 and U2. In method eight, a covariance of 0.000504 was identified, implying a correlation of 0.92. For method nine, 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 error.

[Figure 3 about here]

### Scatterplot comparison

In Figure 4 the scatterplots of U2 against U1 are shown for each of the ten methods. The scatterplot from bootstrapping the IPD, shown in full size in Figure 1, 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, four, five and six 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, three, four, five, and six 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. We see from the scatter that methods eight and ten are closest in appearance to the bootstrapped scatter.

[Figure 4 about here]

### 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.

### Violin plot comparisons of U1, U2 and U1-U2

Figure 5 present the distributions of U1, U2 and U1-U2.The distributions of estimates of U1 and U2 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 distribution of estimates of U1 – U2, i.e. the differences in paired draws of U1 and U2. Method one, independent sampling, producing some estimates which violate monotonicity. 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 an underestimation of the true uncertainty in this quantity, evident most strongly in methods two and seven. The second problem is evident in methods 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 about here]

### Conclusion from first comparison

Based on the above results, methods eight, nine and ten appeared most promising, and were evaluated further using a second hypothetical dataset.

## Second hypothetical dataset, three best methods

In the second hypothetical dataset U1 and U2 are derived from different samples of individuals, and again it is assumed that no IPD are 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 *U*2\*, are no longer very similar. Instead the distribution of *U*2\* is wider than of *U*2.

Figure 6b shows the scatterplot produced by the first variation of the covariance methods, method eight. 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 nine. This is similar to the scatterplot for method eight, 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.

The scatterplot for method ten 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 about here]

# 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 using resampling or replacement(3-6) introduce implausible discontinuities into the distribution of differences between values (the inter-distribution?); 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 4c, only methods eight, nine and ten appeared to be broadly appropriate in their representation of both intra-distribution uncertainty and inter-distribution

However, in theory the multivariate model methods (seven to nine) may all be inappropriate, because of the dependence on the covariance parameters used on the sample sizes of the training sets used to estimate them. As normal distributions are not bounded, for infinitely sized samples the probability that monotonicity becomes violated at least once becomes a certainty, and so the covariance parameter which implies a correlation of 1 always becomes selected. In doing this, either U1 becomes an analytic derivation of U2 or vice-versa. Information about the standard error of the derived parameter therefore becomes disregarded and one One of the few sample summary statistics available to the modeller is therefore discarded

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; it is thusanalytically impossible that method ten will produce any pairs of estimates which violate the monotonicity assumption. Secondly, 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 describes limitations with the current analysis. These include: 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 is 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 monotonicity condition becomes and . Introducing further states would lead to additional complexity for methods seven, eight and nine, for example, as we would have to estimate three covariances (i.e. cov(U0, U2), cov(U0, U1) and cov(U1,U2) ) rather than just one. This provides further justification for favouring 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 relates to the existence of worse-than-death health states exist where negative values are recorded and where it would be inappropriate to use Beta distributions 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 HRQOLHRQOL scores using the EuroQol Quality of Life (EQ-5D) scale. (5,6)

## Implications for Research and Practice

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 reported a higher HRQoL score in the more severe state than in the less severe 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 health states. 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.

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. As described above, there are theoretical reasons to be cautious about the use of methods seven to nine, although in practice estimates produced by these methods may be reasonable.

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

Our comparison between methods suggests the difference method is preferable to the other methods considered. For this reason, we recommend this method be adopted within health technology appraisals where monotonicity between parameters is believed to exist. The R code for implementing all ten methods is included as an appendix, and an Excel worksheet implementing the difference method online.

# 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

# Tables

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

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

|  |  |  |
| --- | --- | --- |
|  | **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 Assumed summary data available in supplementary analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| **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 A1 Summary of the ten approaches considered. The † symbol indicates classes of methods where monotonicity cannot be violated.