

Deriving weights using 20 seed questions

Following the results of the above analysis of deriving weights with 10 seed questions, it is of interest to assess whether more conclusive results could be obtained if we use more seed questions for deriving weights. Hence, two data sets from the data base of complete EJ analysis program (*THRMBLD* and *TNODISPR*) were selected for the analysis. Similar to the case of 10 seed questions above, data contain experts' elicited 5% , 50%, and 95% percentiles of the probability distributions of quantities of some given seed questions in both data sets. First 20 questions from each data set were used to derive weights and the remaining questions were used to estimate variances of weights and to compare the overall calibration and information scores of typical and shrinkage weights as discussed above.

There are some differences between the two data sets. *TNODISPR* data set contain very small realized and elicited percentiles values for most of the seed questions compared to the *THRMBLD* data set. Therefore, it was of interest to assess the impact of the magnitude of the realized and the elicited values on the overall calibration and information scores of the testing questions based on the normalized typical and shrinkage Cooke's weights in the analysis. Table 1 summarises the DM scores for both testing data sets. Overall, the calibration scores are small for both data sets under two types of weights and the information scores are relatively large. However, we could not identify considerable differences between scores as similar to the case of using 10 seeds questions for deriving weights above.

Table S1: Overall Decision Maker scores of testing questions

Dataset	Types of weight	Calibration score	Information score
THRMBLD	Typical	0.007814	0.4112
	Shrinkage	0.007814	0.3963
TNODISPR	Typical	0.05176	0.7012
	Shrinkage	0.05176	0.6826