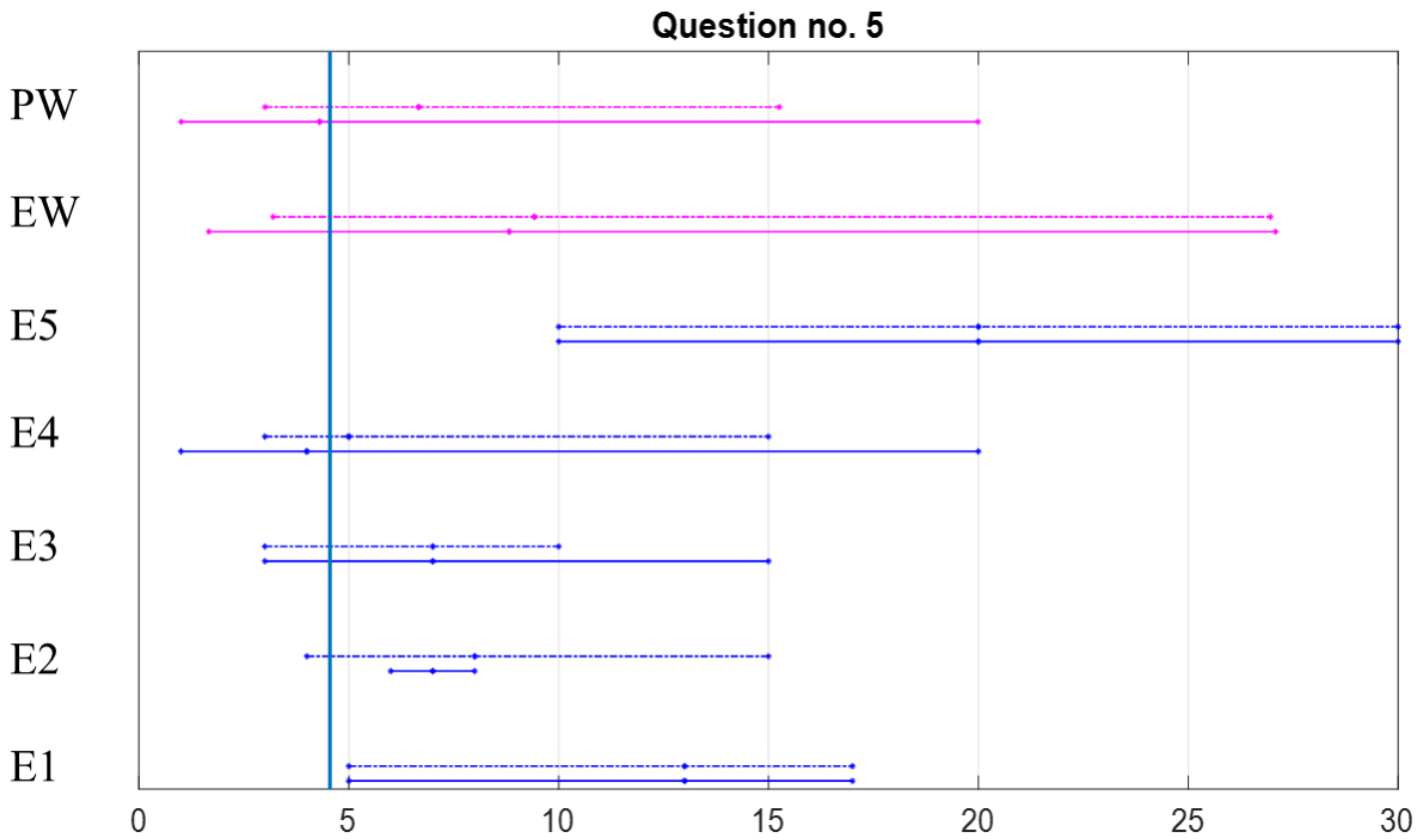
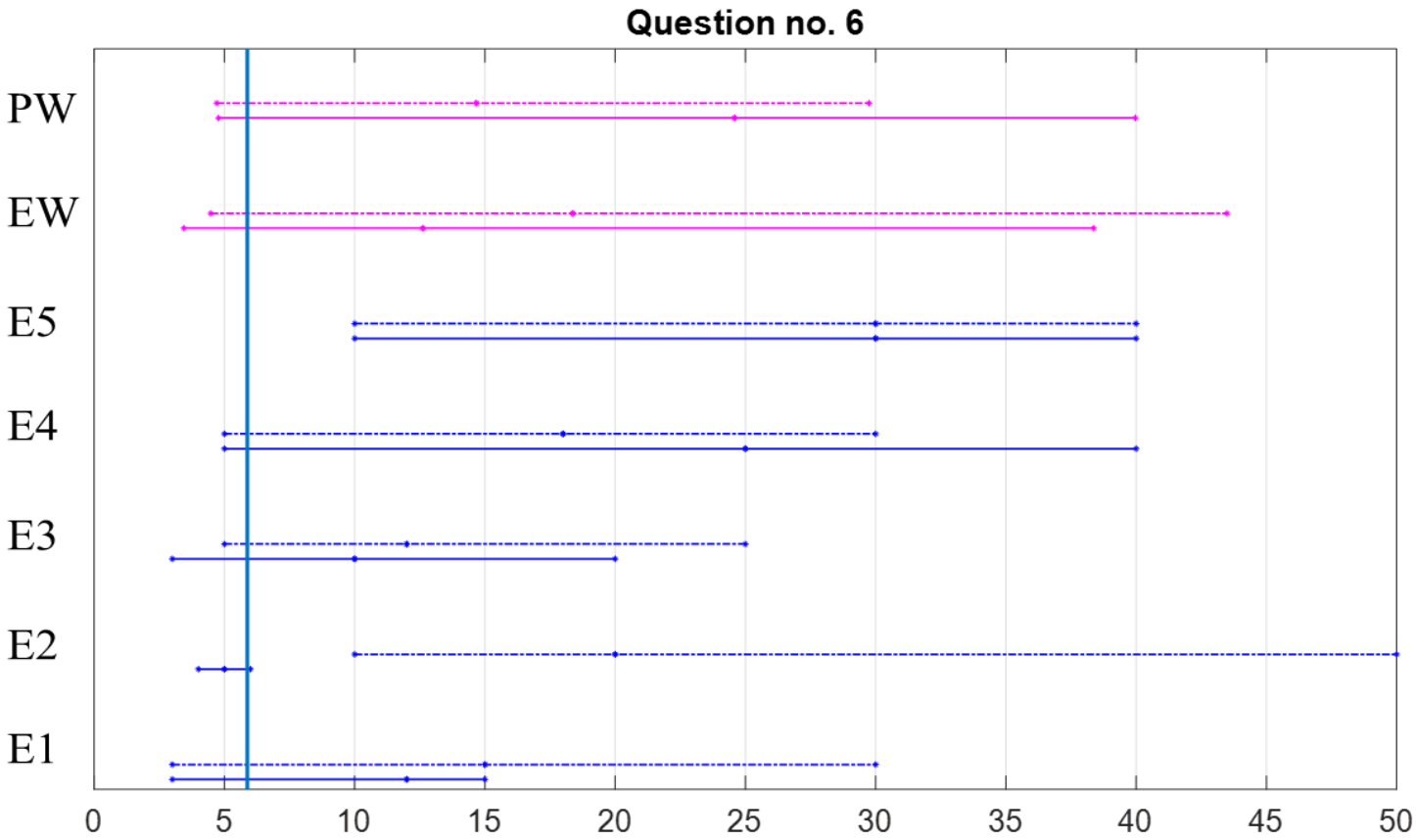
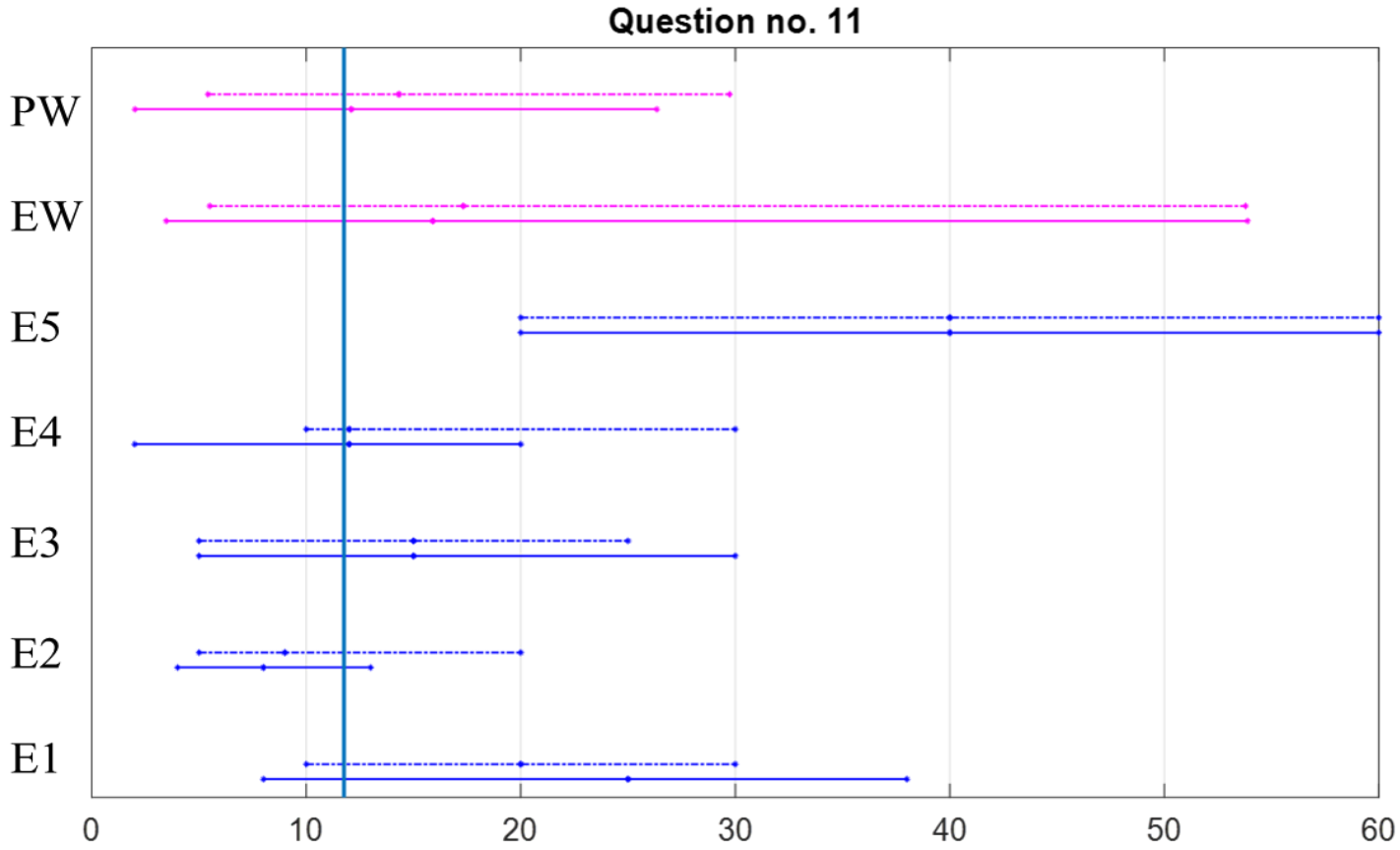


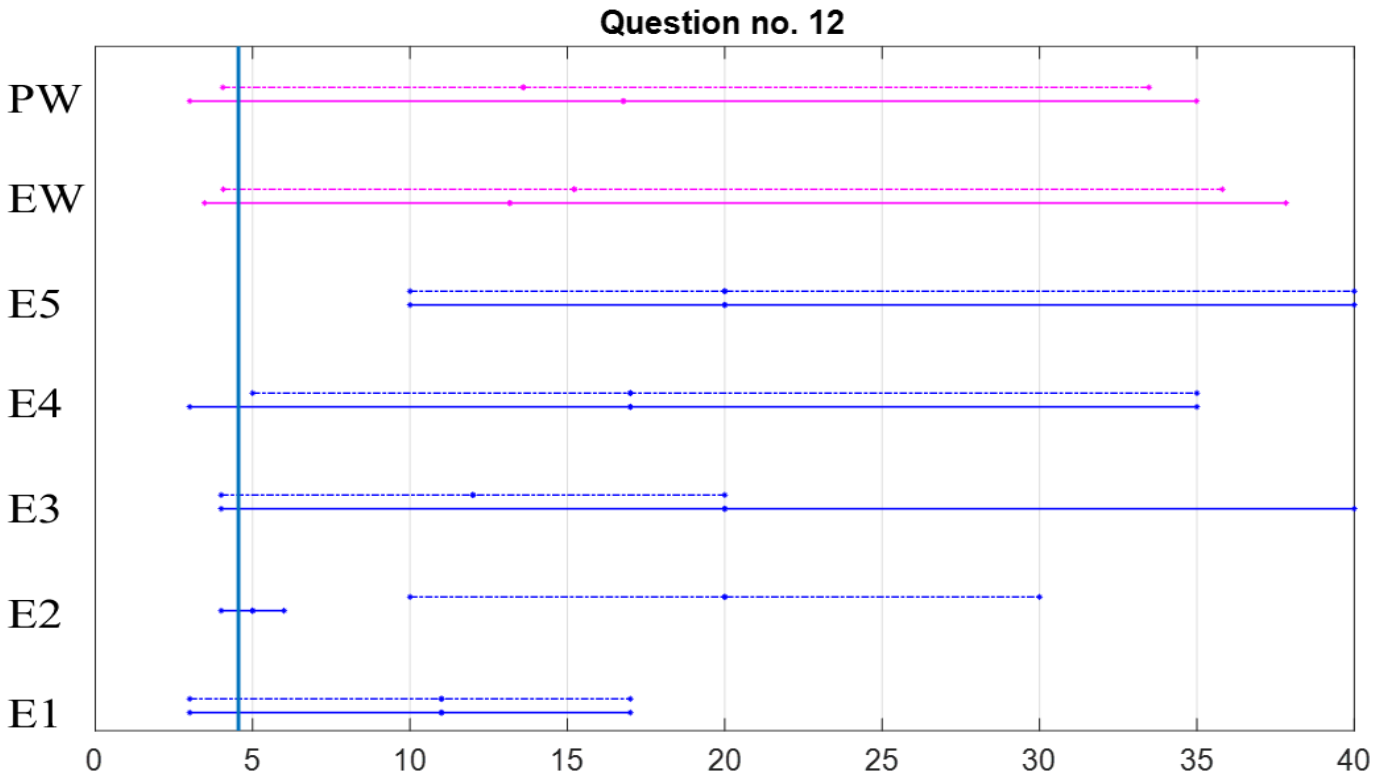
Supplementary Figure 13 Calibration Q5: In both cases the pooled Round 2 (R2) estimates were further from the truth than the first round, but not by as much in the performance-weighted (PW) (top) line. Numbering from the bottom, Expert 2 (E2) is very confident in the first round (narrow interval between 5th & 95th estimate) but less so in the second round. In R2, their best estimate moves further from the truth but their intervals now encapsulate the truth, as do those of all the other experts but Expert 5 (E5). Experts 2 (E2), 3 (E3) & 4 (E4) all adjust their estimates in R2, and 3 & 4 reduce their uncertainty. Expert 1 (E1) & 5 (E5) give identical R1 and R2 estimates. Expert 5 does this throughout.



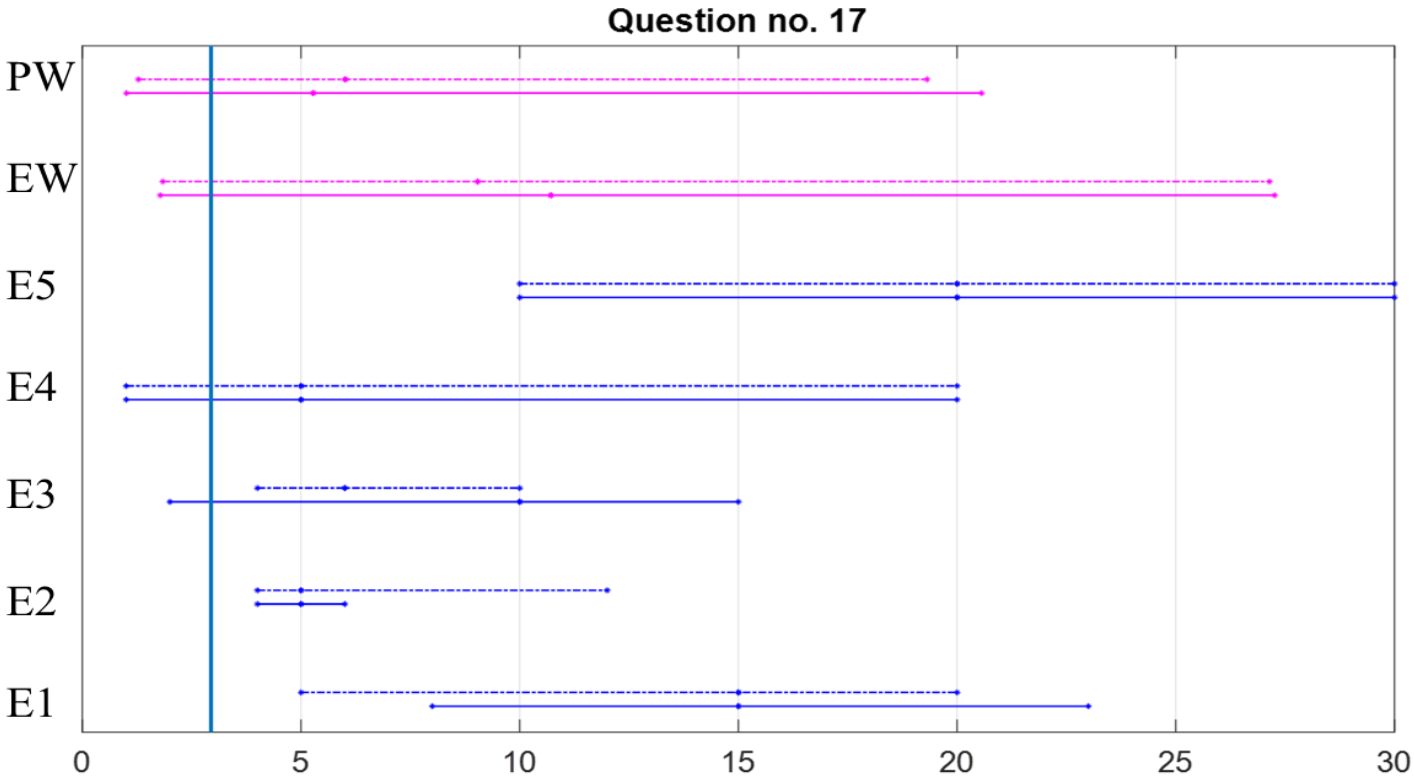
Supplementary Figure 14 Calibration Q6: In Round 2 the performance-weighted (PW) pooled estimate moves strongly towards the truth whilst the equal-weighted (EW) pooled estimate moves further away. All experts capture the truth within their intervals in Round 1, but Expert 2 (E2) moves much further away in Round 2 and increases their intervals very significantly. Expert 4 (E4) is the only one to move in the direction of the truth in Round 2; Expert 5 (E5) is unchanged by discussion, and their estimate fails to encapsulate the truth despite very wide uncertainty bands.



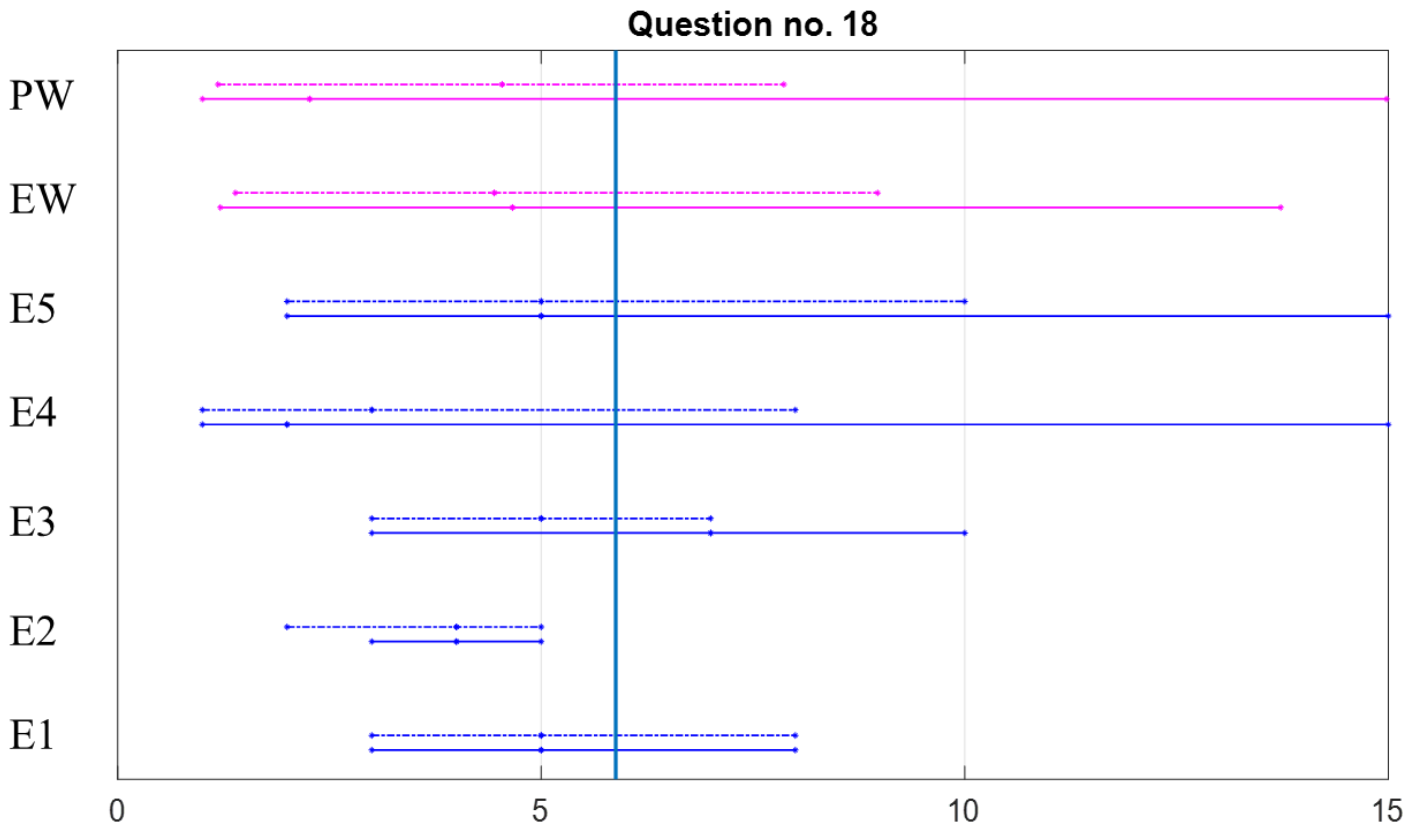
Supplementary Figure 15 Calibration Q11: Both the pooled Round 2 estimates are further from the truth than the Round 1 estimates, but the performance-weighted estimates are both closer to the truth than the equal weighted estimates. Expert 4 (E4) produces the same very accurate estimate in both rounds, but shifts their uncertainty interval upwards following the discussion. Expert 1 (E1) & 2 (E2) move towards the truth in Round 2, Expert 3 (E3) does not change their best estimate but narrows their intervals. Expert 5 (E5) is unchanged by discussion, and their estimate fails to encapsulate the truth despite very wide uncertainty bands.



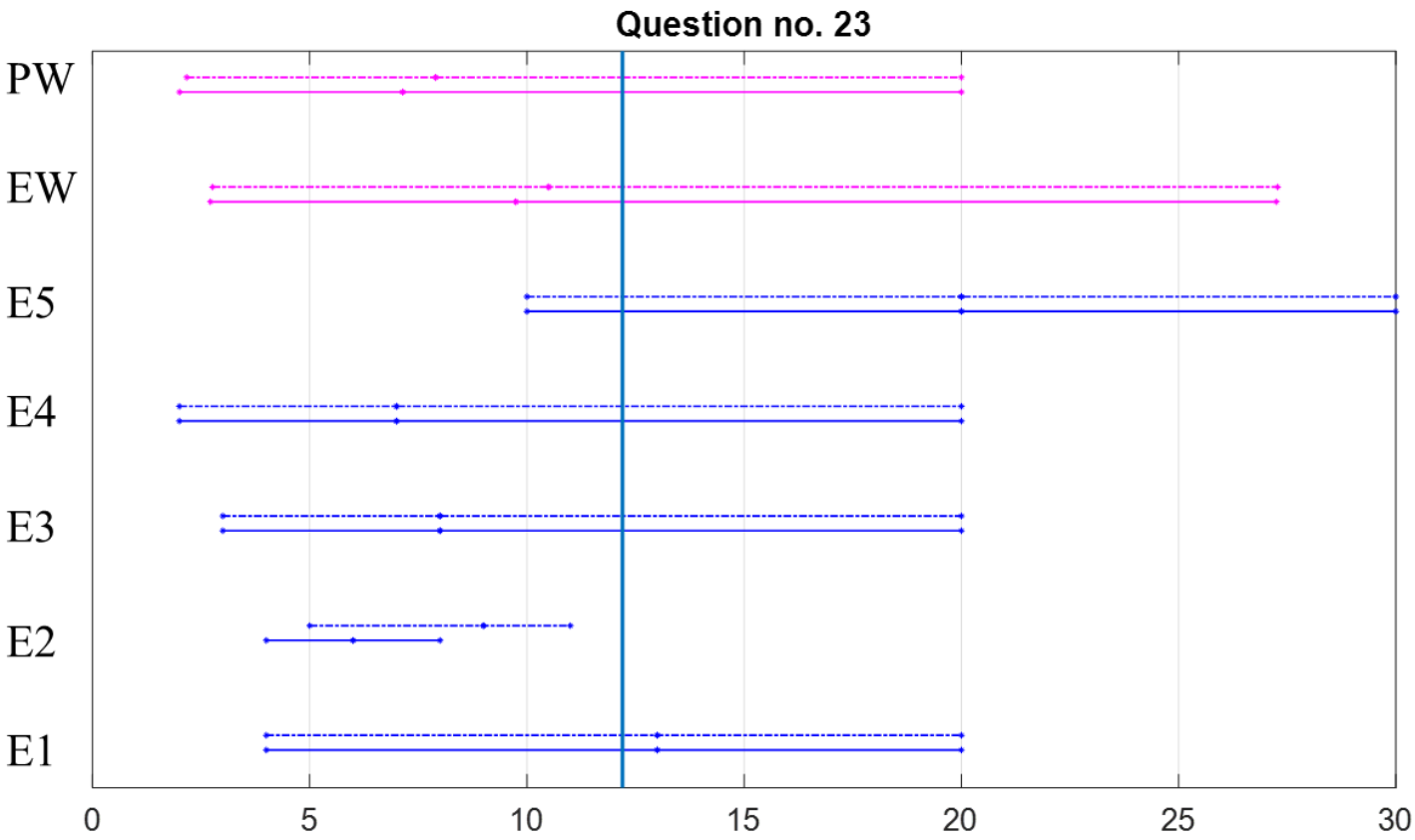
Supplementary Figure 16 Calibration Q12: The performance-weighted (PW) pooled estimate moves toward the truth in Round 2 whilst the equal weighted (EW) pooled moves further away. Expert 2 (E2) gives a good estimate with narrow uncertainty bands (the difference between the highest plausible and lowest plausible values) in Round 1 but their Round 2 estimate has much wider intervals, does not overlap their Round 1 estimate and is much further from the truth. Expert 3 (E3) moves much closer to the truth in Round 2 with narrower bands. Expert 4 (E4) slightly reduced their uncertainty in Round 2. Expert 1 (E1) is unchanged in Round 2 but does capture the truth in their uncertainty bands. Expert 5 (E5) is unchanged by discussion, and their estimate fails to encapsulate the truth despite very wide uncertainty bands.



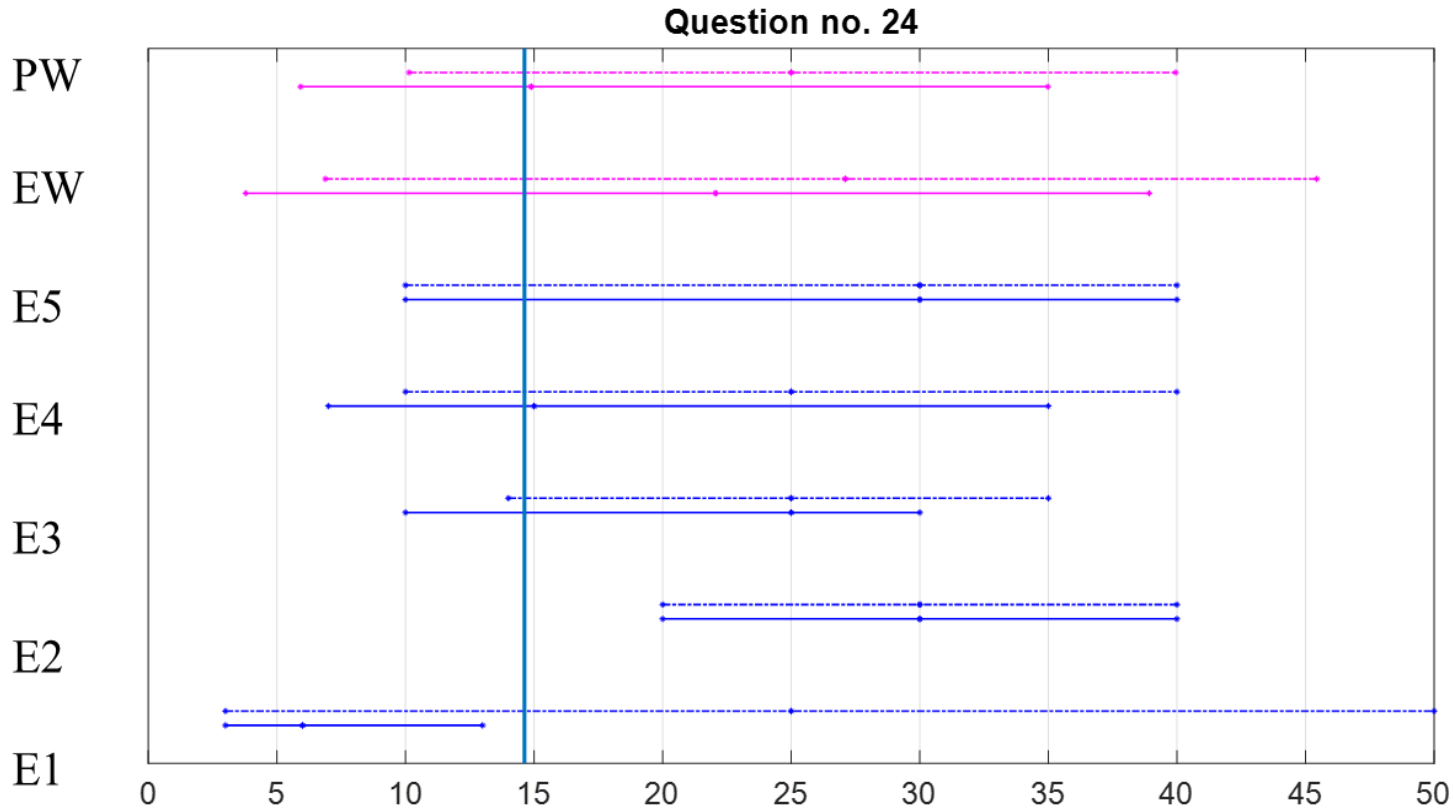
Supplementary Figure 17 Calibration Q17: The performance-weighted (PW) pooled estimate moved further from the truth in Round 2, but only slightly, and is much nearer than either the Round 1 or Round 2 estimates for equal weighted (EW) pooled estimate. In Round 2, only Expert 4 (E4) captured the truth within their uncertainty bands (the difference between the highest plausible and lowest plausible values), although Expert 2 (E2) gave the same best estimate. Expert 5 (E5) is unchanged by discussion, and their estimate fails to encapsulate the truth despite very wide uncertainty bands.



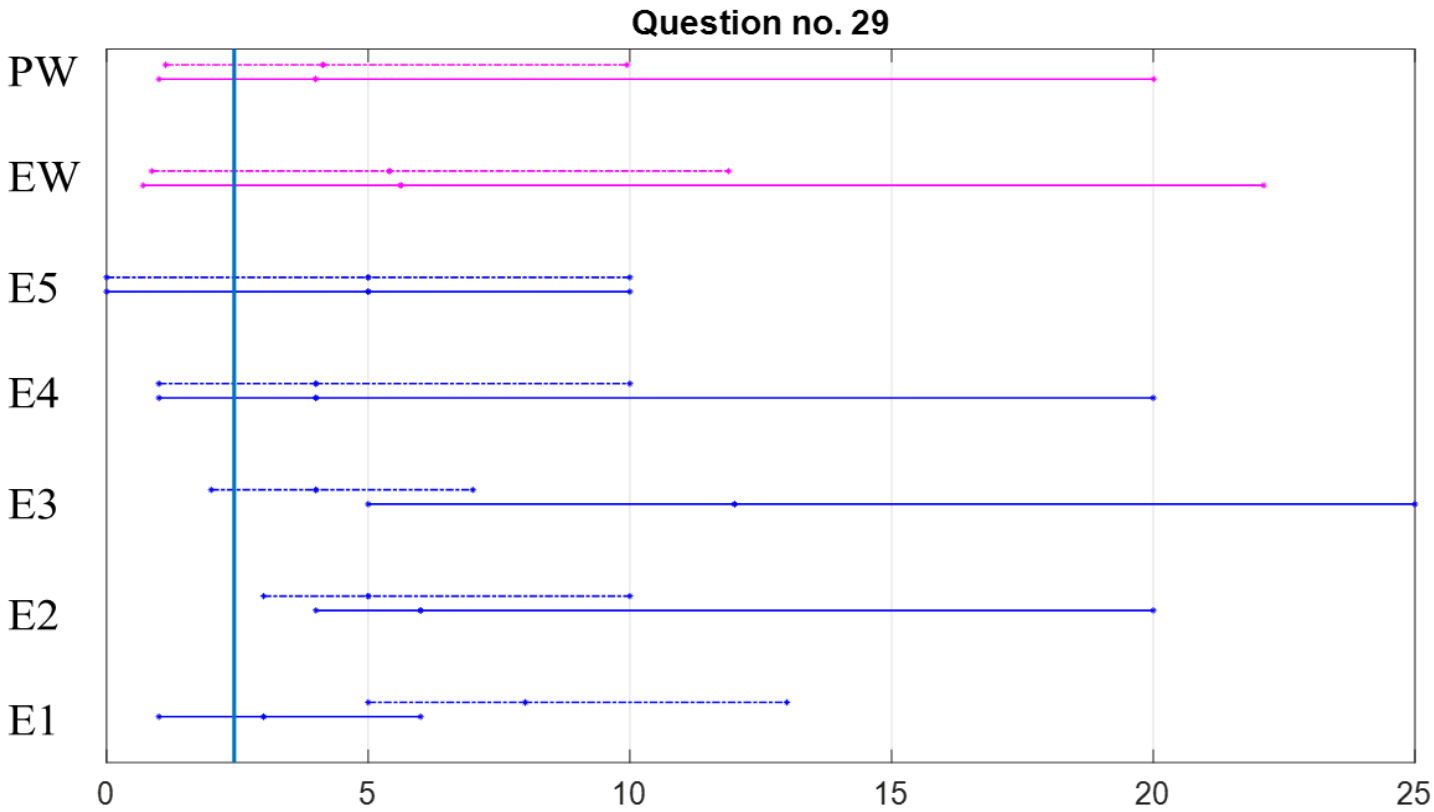
Supplementary Figure 18 Calibration Q18: The performance-weighted (PW) pooled estimate moved strongly in the direction of the truth in Round 2, whilst the equal weighted (EW) moved slightly away, but giving similar estimates. In Round 2 three experts gave the same estimate, which was close to the truth. Only Expert 2 (E2) failed to encapsulate the truth within their uncertainty bands (the difference between the highest plausible and lowest plausible values). Expert 3 (E3), 4 (E4) & 5 (E5) narrow their uncertainty bands after discussion, and Expert 5's estimate, unusually, encapsulates the truth.



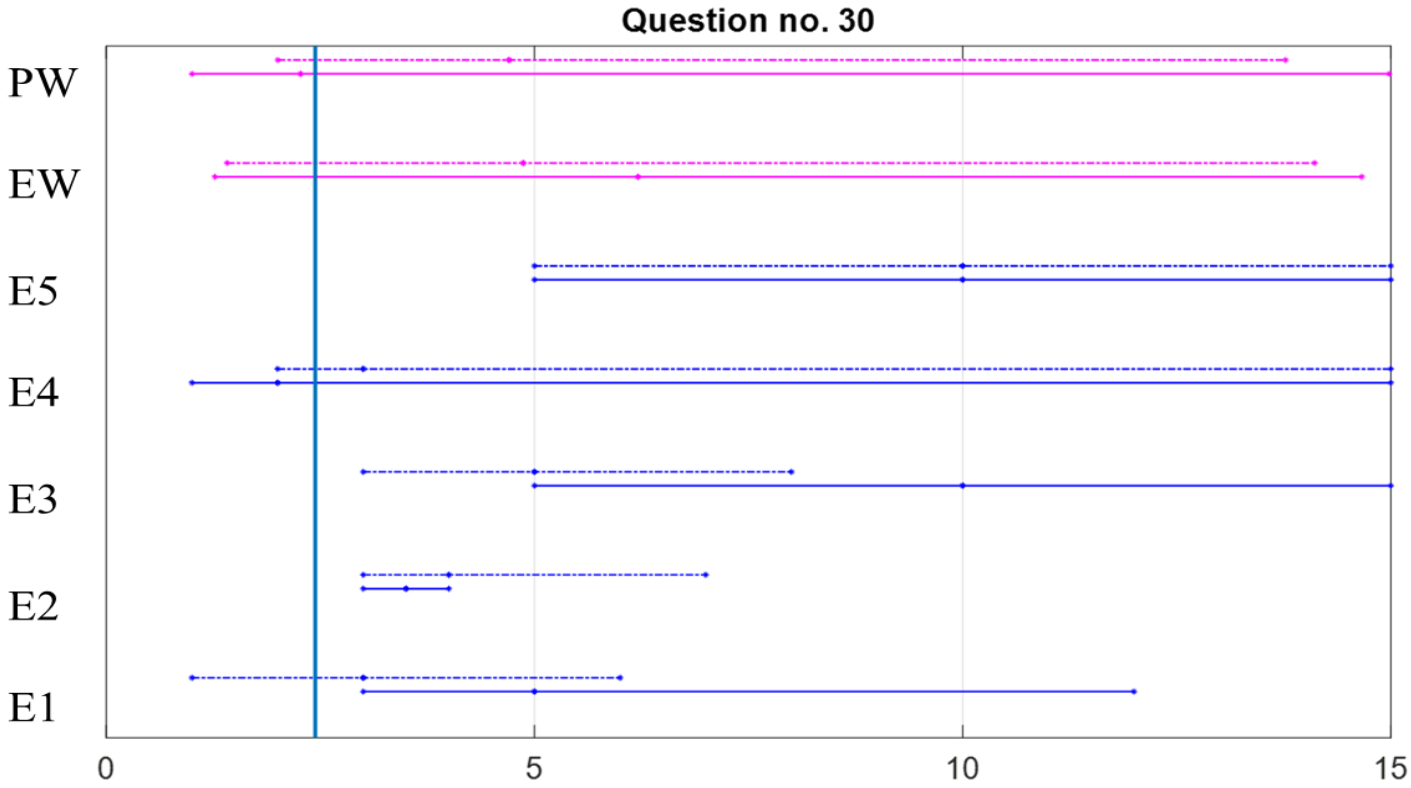
Supplementary Figure 19 Calibration Q23: performance-weighted (PW) pooled estimate was further from the truth than equal weighted (EW) pooled estimate, but moved in the direction of the truth in Round 2. The only expert that changes their estimate in Round 2 was Expert 2 (E2), who moved toward the truth with wider uncertainty bands (the difference between the highest plausible and lowest plausible values). Expert 5 (E5) is unchanged by discussion, but their estimate does just encapsulate the truth in very wide uncertainty bands.



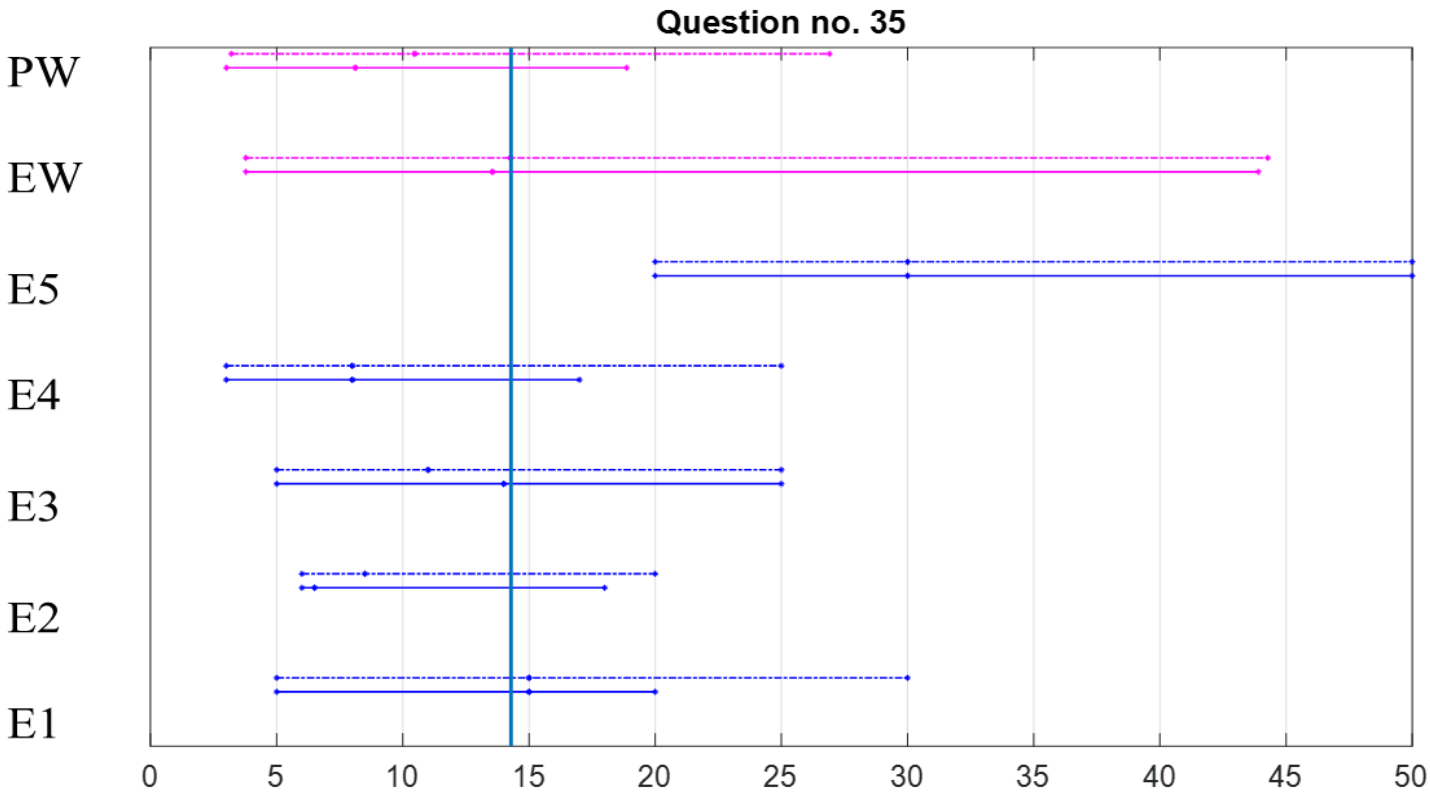
Supplementary Figure 20 Calibration Q24: Both Performance weighted (PW) & equal weighted (EW) pooled estimates moved further from the truth in Round 2, but performance-weighted is more accurate in both rounds. Expert 3 (E3) retains the same best estimate with shifted uncertainty bands. Expert 1 (E1) moves the most making an overestimate in Round 2 which is further from the truth than their underestimate in Round 1. Expert 2 (E2) is unchanged by discussion, and their estimate fails to encapsulate the truth, and Expert 5 (E5) is unchanged in Round 2 but their estimate does just encapsulate the truth in very wide uncertainty bands (the difference between the highest plausible and lowest plausible values).



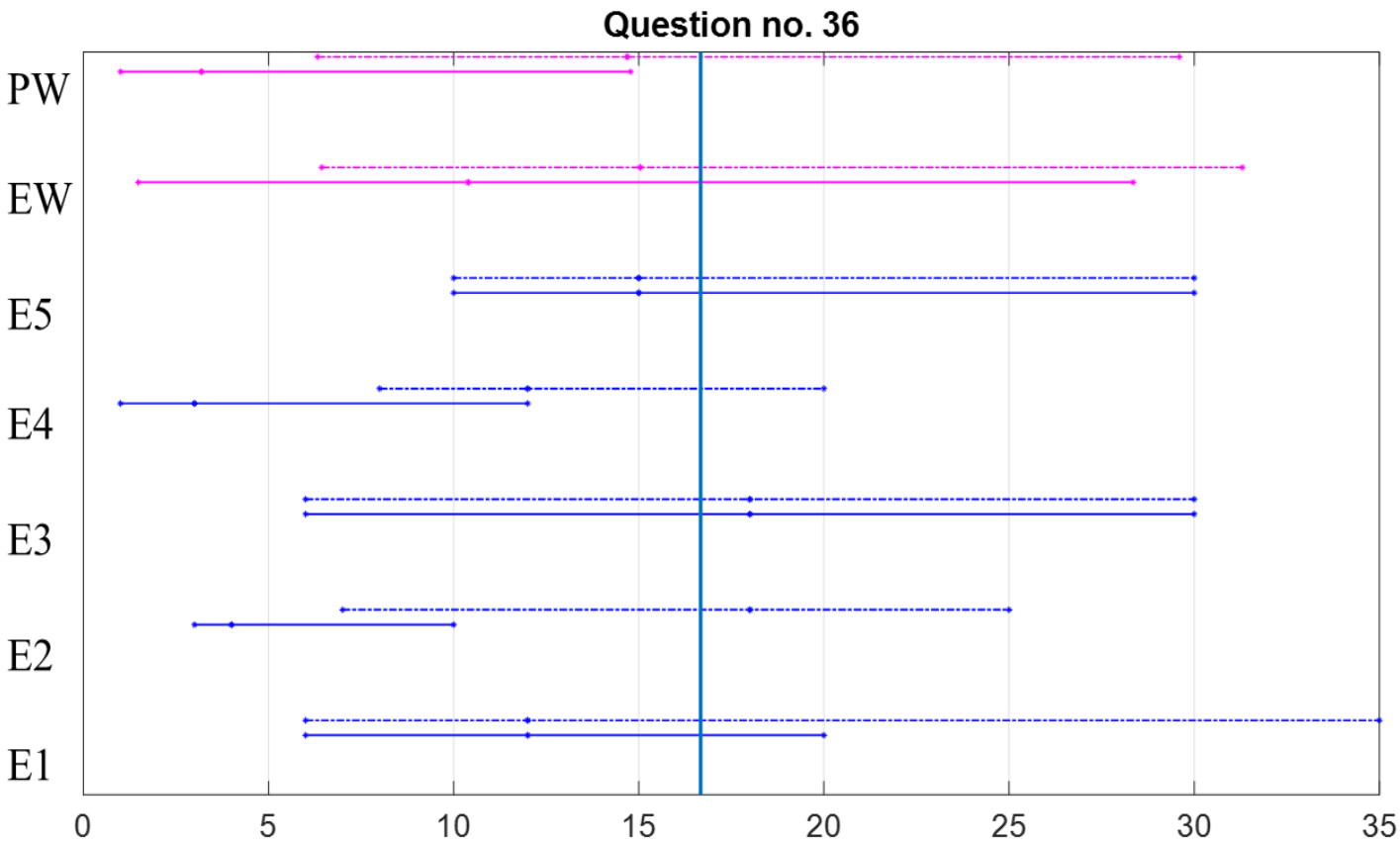
Supplementary Figure 21 Calibration Q29: Performance-weighted (PW) pooled estimate provided a better estimate than equal weighted (EW) despite shifting slightly away from the truth in Round 2. Experts 2 (E2) and 3 (E3) moved towards the truth in Round 2 and reduced the width of their uncertainty bands, but Expert 2 failed to encapsulate the truth. Expert 1 (E1) moved away from the truth with wider uncertainty bands in Round 2. Expert 4 (E4) retained the same best estimate with narrower uncertainty bands (the difference between the highest plausible and lowest plausible values) in Round 2. Expert 5 (E5) is unchanged in Round 2 but their estimate does encapsulate the truth in very wide uncertainty bands.



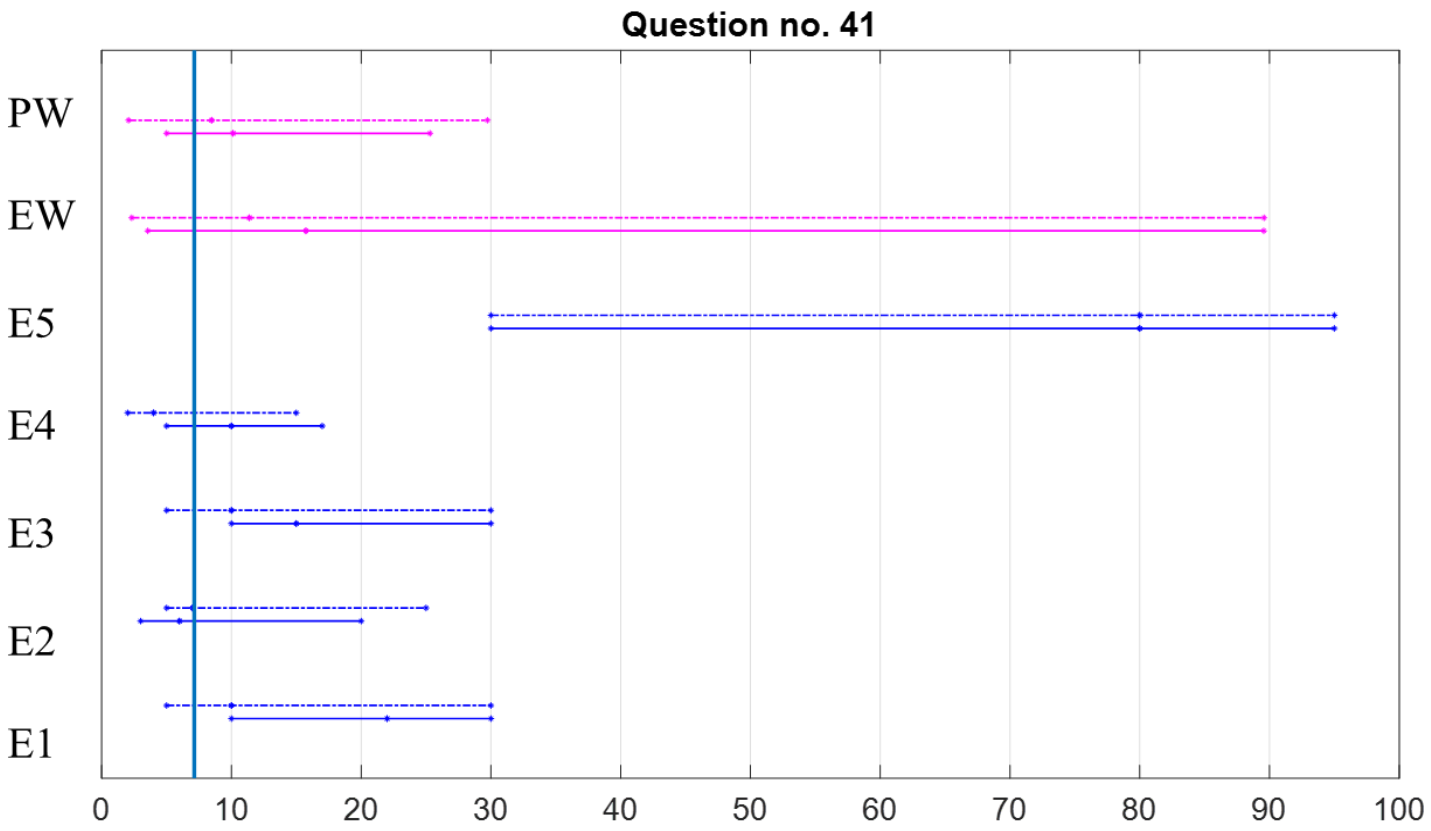
Supplementary Figure 22 Calibration Q30: Performance weighted (PW) & equal weighted (EW) pooled estimates provide equally good Round 2 estimates, with performance-weighted pooled estimate moving away from the truth and equal weighted toward it in Round 2. Only Expert 1 (E1) & 4 (E4) encapsulate the truth in their uncertainty bands (the difference between the highest plausible and lowest plausible values). Experts 1 (E1) & 3 (E3) move towards the truth in Round 2 whilst Expert 2 & 4 move away from it. Expert 5 (E5) is unchanged by discussion, and their estimate fails to encapsulate the truth despite very wide uncertainty bands.



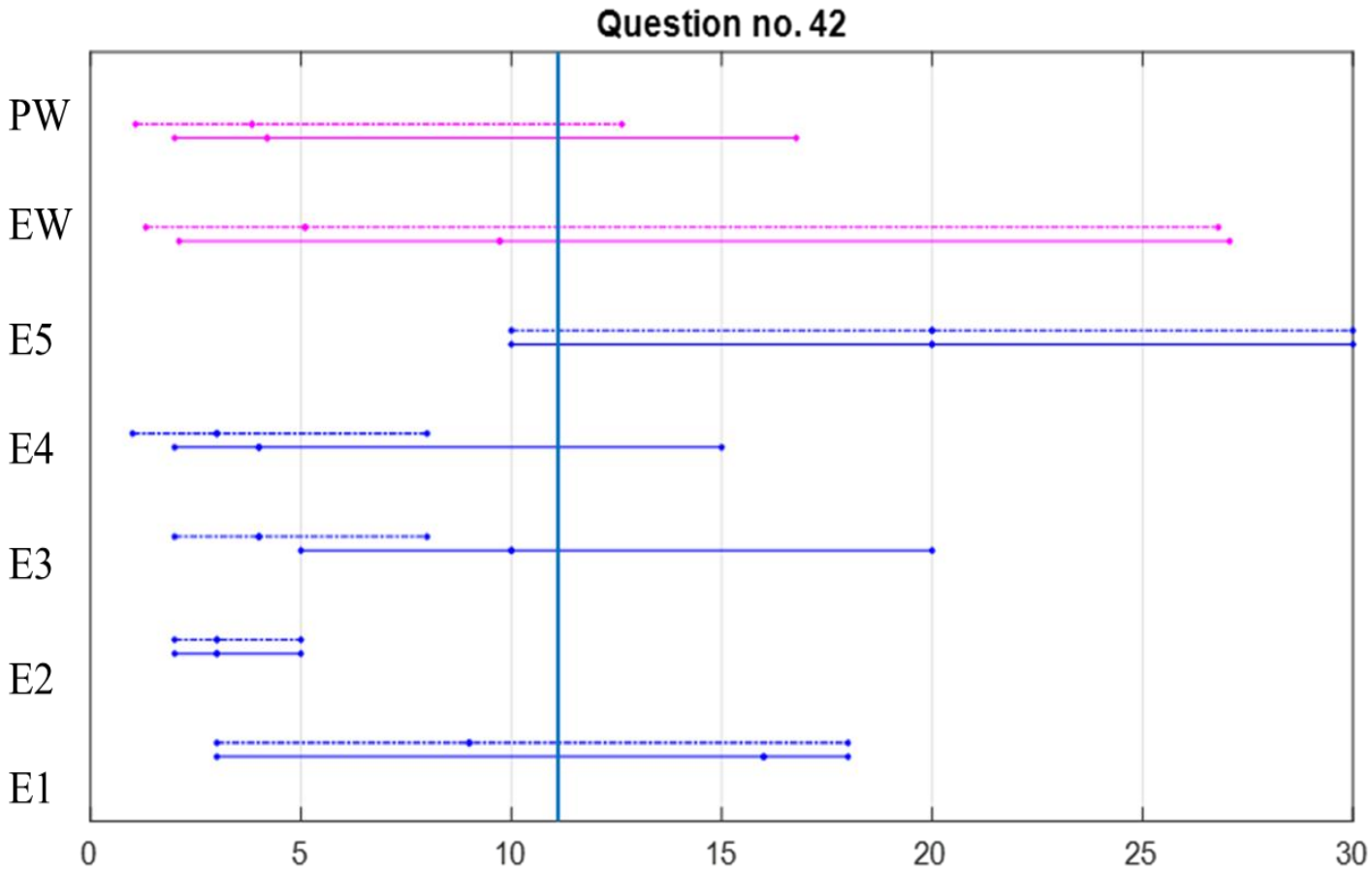
Supplementary Figure 23 Calibration Q35: Equal weighted (EW) pooled estimate gives a better estimate than performance-weighted (PW) in both rounds, with performance-weighted moving towards the truth in Round 2. This is because Expert 5's (E5) overestimate is balanced by underestimates from other experts. Expert 2 (E2) moves towards the truth and Expert 3 (E3) moves away from it. Experts 1 (E1) and 4 (E4) do not change their best estimates, but adjust their uncertainty bands (the difference between the highest plausible and lowest plausible values) in opposite directions. Expert 5 is unchanged by discussion, and their estimate fails to encapsulate the truth despite very wide uncertainty bands.



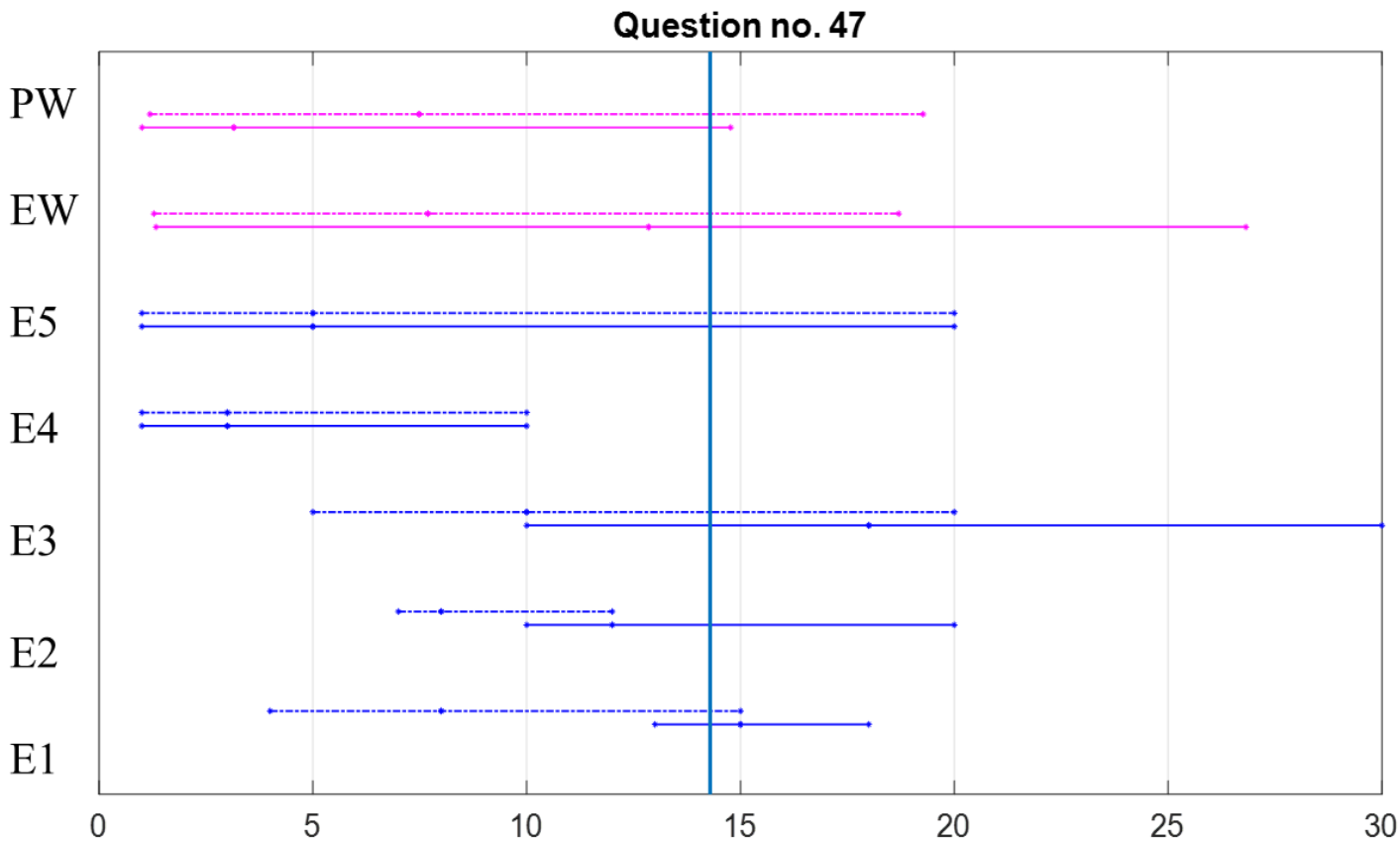
Supplementary Figure 24 Calibration Q36: Equal weighted (EW) pooled estimate provides better Round 1 & Round pooled estimates, both Performance weighted (PW) & equal weighted moving towards the truth in Round 2. Experts 2 (E2) and 4 (E4) fail to encapsulate the truth in their uncertainty bands (the difference between the highest plausible and lowest plausible values) in Round 1 but move towards the truth and encapsulate it in Round 2, All other experts do not change their best estimate, although Expert 1 (E1) does increase their uncertainty bands. Unusually, Expert 5 (E5) provides a good estimate, albeit with very wide uncertainty bands.



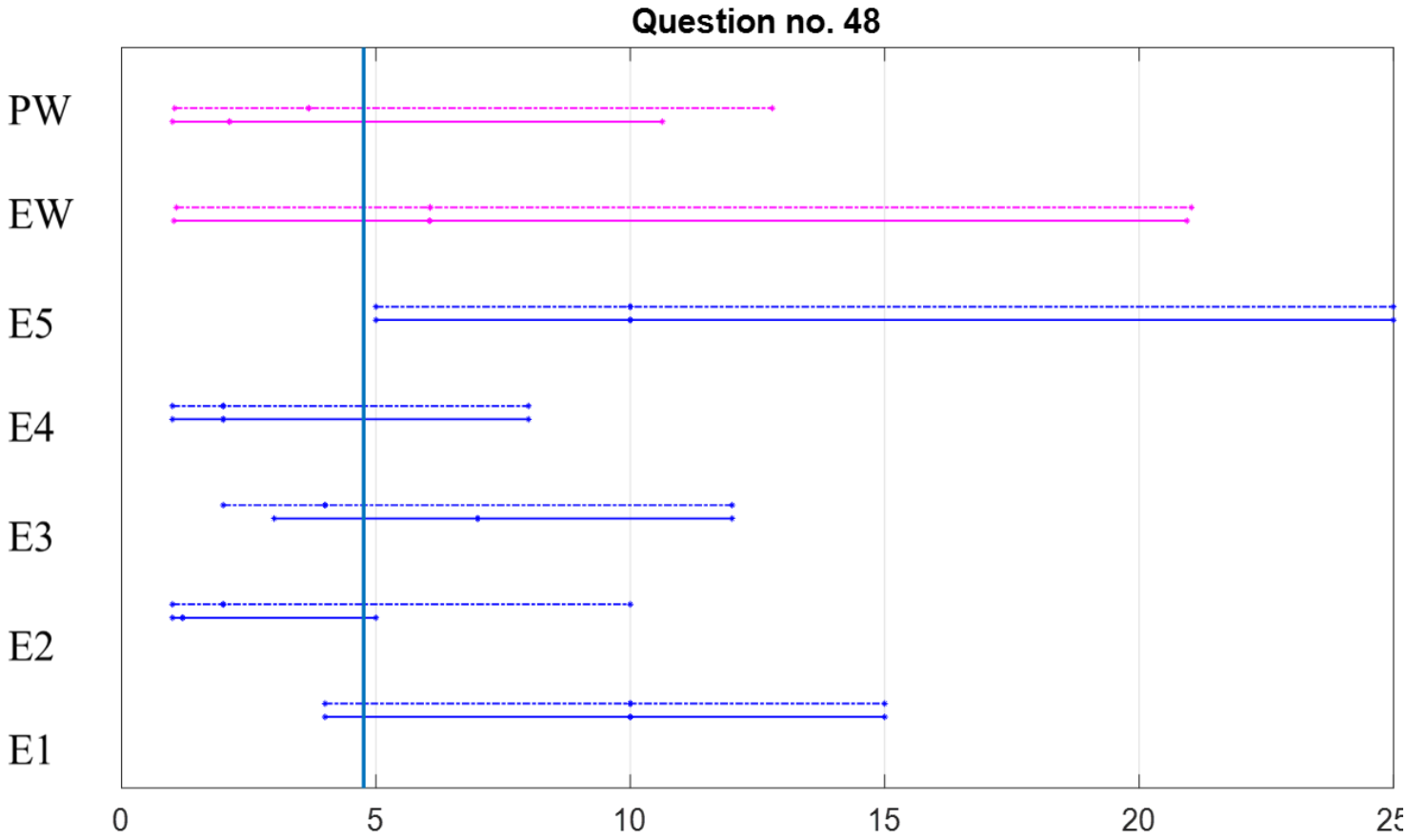
Supplementary Figure 25 Calibration Q41: Performance weighted (PW) & equal weighted (EW) pooled estimates provide good estimates with performance-weighted closer to the truth and having narrower uncertainty bands (the difference between the highest plausible and lowest plausible values). All experts except 5 (E5) move towards the truth and provide good estimates in Round 2. Experts 1 (E1) & 3 (E3) encapsulate the truth in Round 2 only. Expert 5 provides extraordinarily poor estimates, unchanged by discussion, and their estimate fails to encapsulate the truth despite very wide uncertainty bands.



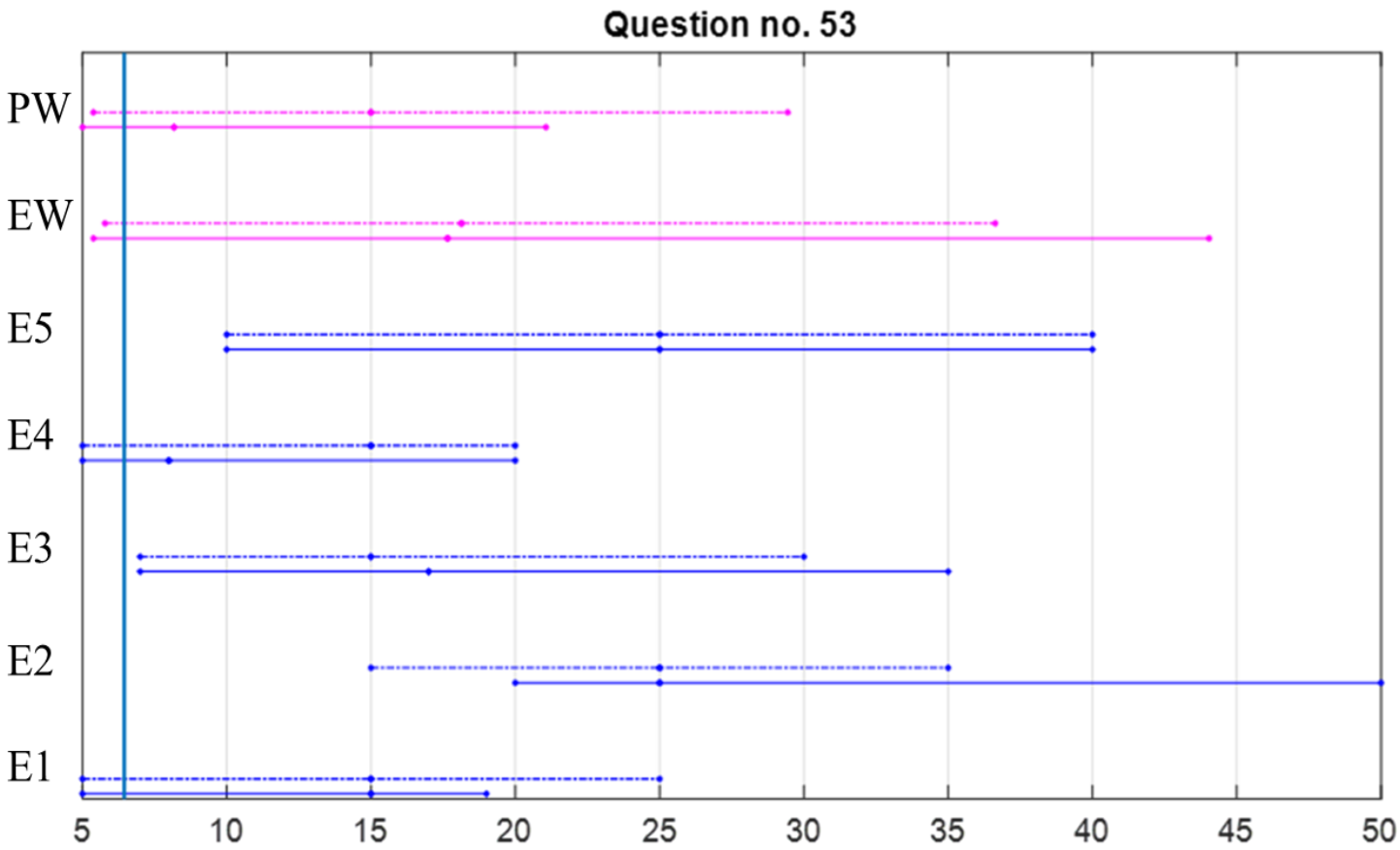
Supplementary Figure 26 Calibration Q42: Equal weighted (EW) pooled estimate gives a better estimate than performance-weighted (PW), both moving away from the truth in Round 2. Experts 3 (E3) & 4 (E4) move away from the truth in Round 2. Expert 1 (E1) retains the same uncertainty bands, which encapsulate the truth, but improves their best estimate in Round 2. Expert 5 (E5) is unchanged by discussion, but their estimate encapsulates the truth within very wide uncertainty bands.



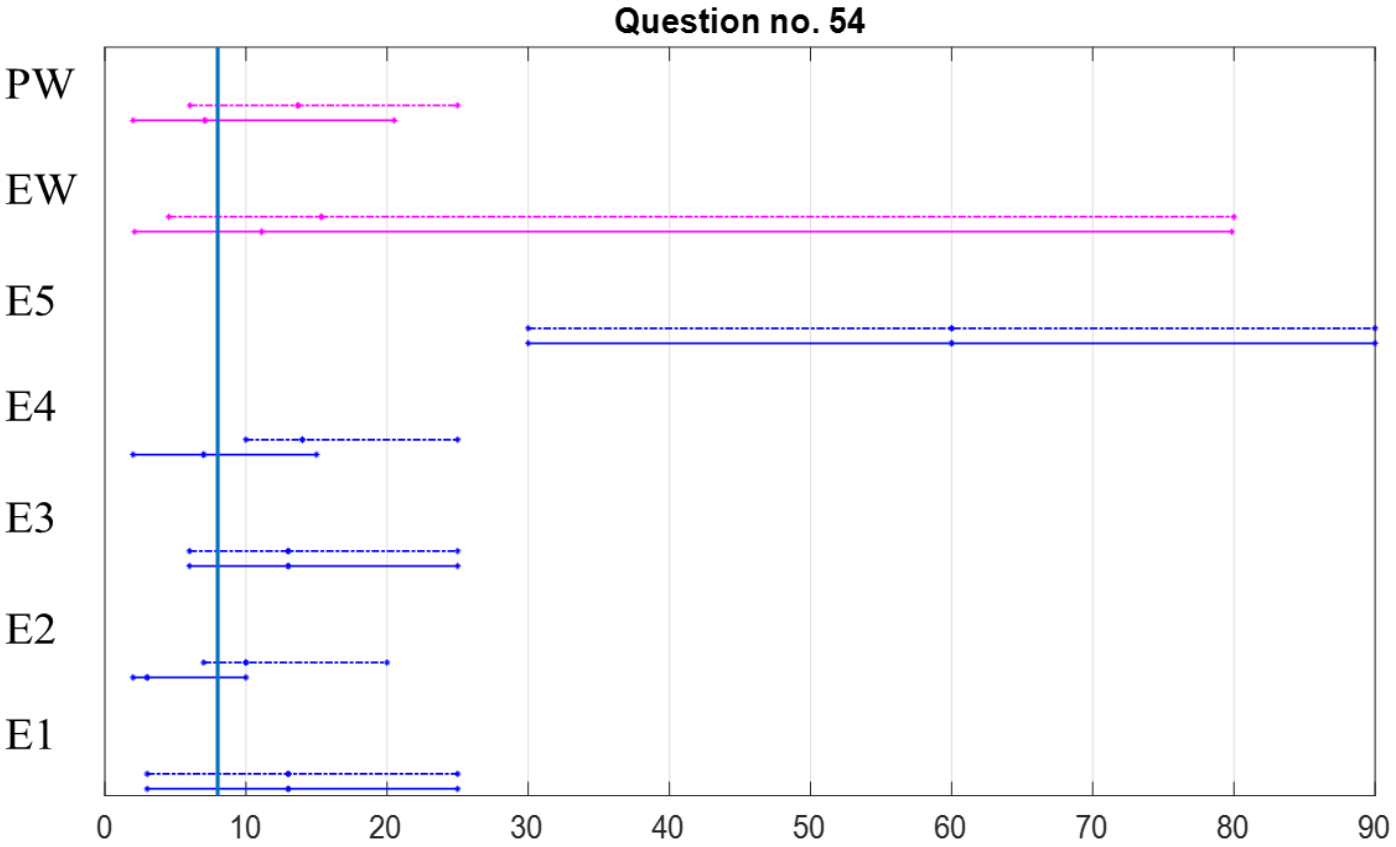
Supplementary Figure 27 Calibration Q47: Performance weighted (PW) & equal weighted (EW) pooled estimates give similar pooled estimates in Round 2 with performance-weighted moving towards the truth and equal weighted away from it. All experts but 4 (E4) encapsulate the truth within their uncertainty bands (the difference between the highest plausible and lowest plausible values), although Expert 2 (E2) does not encapsulate the truth in Round 2. Experts, 4 & 5 (E5) are unchanged in Round 2, whilst Experts 1, 2 & 3 moving away from the truth in Round 2.



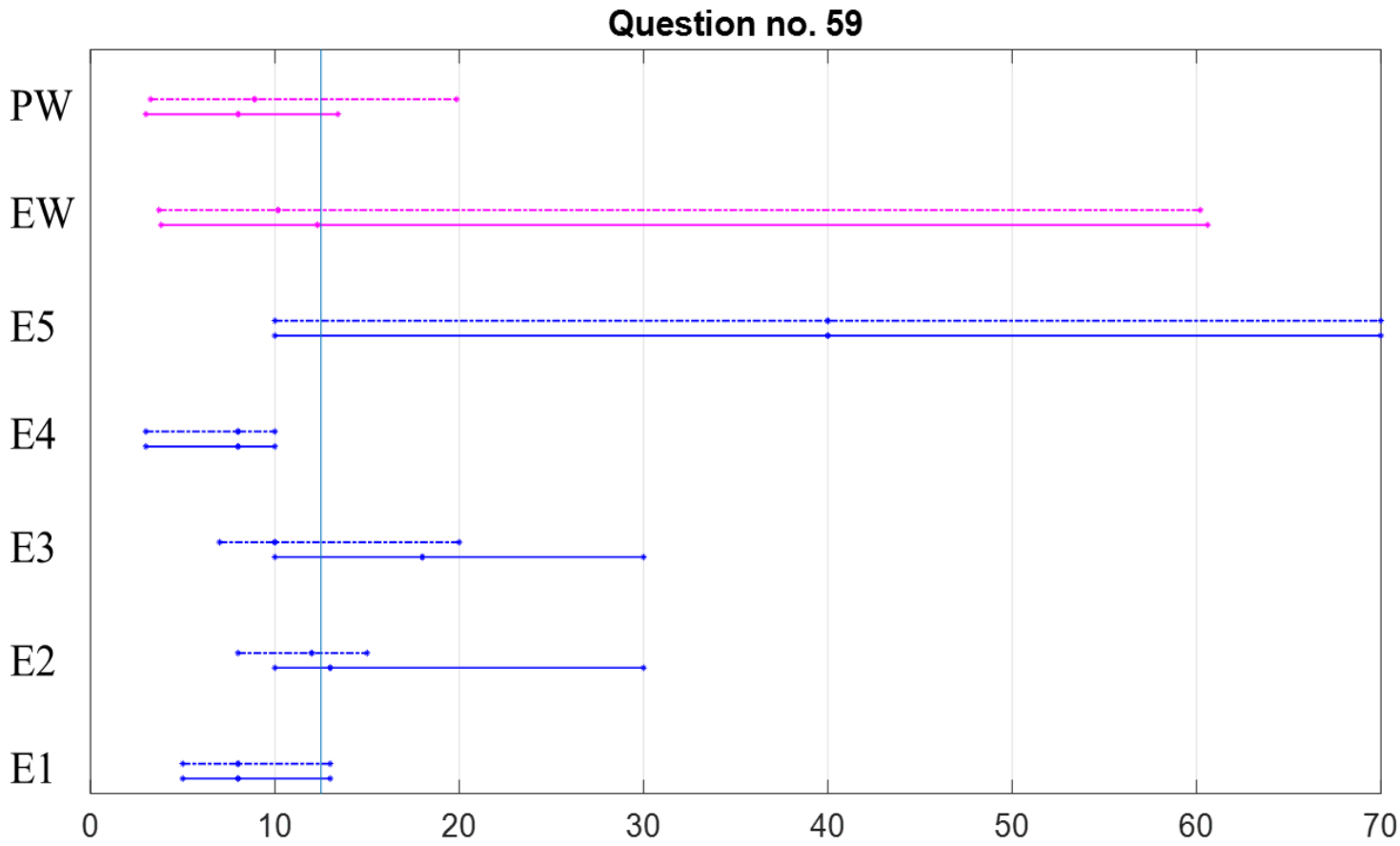
Supplementary Figure 28 Calibration Q48: Both Performance weighted (PW) & equal weighted (EW) give good pooled estimates, with performance-weighted moving towards the truth in Round 2. Experts 2 (E2) & 3 (E3) move towards the truth in Round 2, whilst the others are unchanged. Expert 5 is unchanged by discussion, and their estimate fails to encapsulate the truth despite very wide uncertainty bands (the difference between the highest plausible and lowest plausible values).



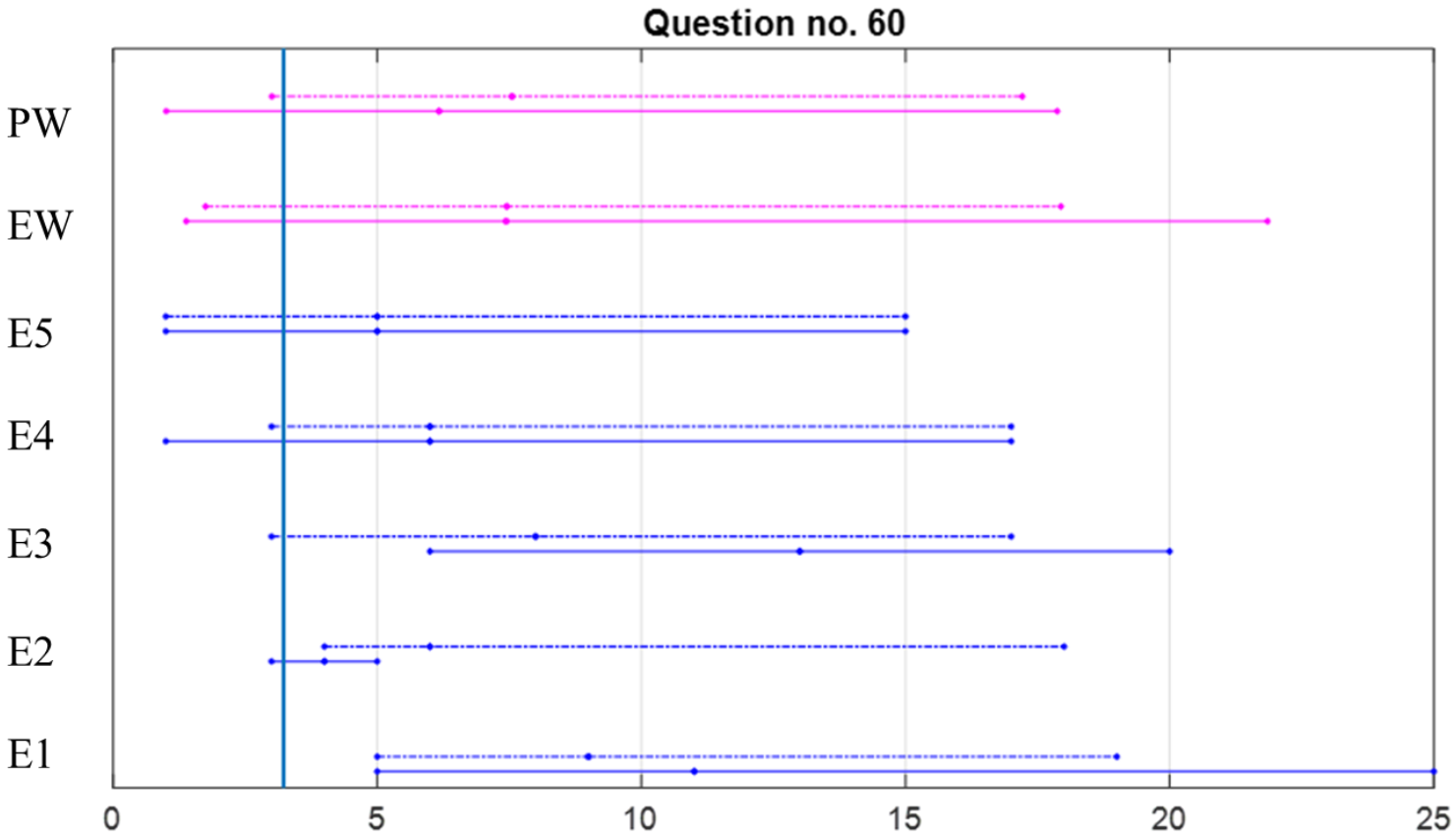
Supplementary Figure 29 Calibration Q53: Performance-weighted (PW) gives a better pooled estimate than equal weighted (EW), despite moving away from the truth in Round 2. Expert 3 (E3) moved towards the truth in Round 2 whilst Expert 4 (E4) moves away from it. The other experts do not change their best estimates, although Expert 1 (E1) and 2 (E2) do change their uncertainty bands (the difference between the highest plausible and lowest plausible values), Expert 2 reducing, Expert 1 increasing. Expert 5 is unchanged by discussion, and their estimate and those of Experts 2 & 3 fail to encapsulate the truth despite very wide uncertainty bands.



Supplementary Figure 30 Calibration Q54: Performance weighted (PW) & equal weighted (EW) pooled estimates provide equally good pooled estimates in Round 2, both moving away from the truth relative to Round 1 where performance-weighted was very accurate. Experts 2 (E2) and 4 (E4) changed their estimate in Round 2, Expert 2 towards the truth and Expert 4 away from it and failing to encapsulate the truth within their uncertainty bands (the difference between the highest plausible and lowest plausible values) in Round 2. Expert 5 (E5) is unchanged by discussion, and their estimate fails to encapsulate the truth despite very wide uncertainty bands.



Supplementary Figure 31 Calibration Q59: Equal weighted (EW) pooled estimate provides a better pooled estimate in Round 1 and Round 2, although the performance-weighted (PW) pooled estimate is close to the truth. Experts 2 (E3) & 3 (E3) change their estimated in Round 2, both towards the truth. Experts 5 (E5), 4 (E4) & 1 (E1) are unchanged by discussion, and Expert 4's estimate fails to encapsulate the truth. Expert 5 just encapsulates the truth using very wide uncertainty bands.



Supplementary Figure 32 Calibration Q60: Performance weighted (PW) & equal weighted (EW) pooled estimates give equally good estimates in Round 2, with equal weighted remaining static and performance-weighted moving towards the truth. Experts 1 (E1) & 3 (E3) move towards the truth in Round 2 whilst Expert 2 (E2) moves away from it. Expert 4 (E4) retains their best estimate but reduces their uncertainty bands (the difference between the highest plausible and lowest plausible values) in Round 2. Expert 5 (E5) is unchanged by discussion, and their estimate encapsulates the truth using very wide uncertainty bands.