Appendix B3. Assessment of Risk of Bias and Small Study Effect

Assessment of Risk of Bias

We slightly modified the Revised Cochrane risk-of-bias tool for randomized trials (RoB 2). We omitted the domain for analysis since that is not relevant for this individual patient data meta-analysis, where we are not reliant on reported trial results. The table below shows scores (1= low risk; 2= some concerns; 3= high risk) for each of the domains and for the overall assessment. The '+' indicates a slightly higher level of concern for bias. Two investigators (DMK and DET) rated all items. Disagreements were resolved by consensus. The risk of bias in the overall assessment reflects the weakest domain.

Appendix Table 6. Risk of Bias Assessment.

Study	Validity Domain				
	Randomization/ Allocation Concealment	Deviations from Intended Intervention (Evidence of large/differential cross-over for 1 treatment)	Bias from Missingness of Outcome Data (<10%; non-differential)	Bias in Outcome Measurement	Overall Assessment
CLOSURE	1	1+	1	2	2
PC Trial	1	1+	2	2+	2+
RESPECT	1	1+	2+	1+	2+
REDUCE	1	1	2	2	2
CLOSE	1	1+	1	2	2
DEFENSE	1	1+	1	2+	2+

Deviations from intended intervention were scored higher when there was large/differential crossover that might reflect patient preference these studies, which were not blinded. Five out of six trials were based on a prospective randomized open blinded end-point (PROBE) design. Since these trials have risk from 'referral bias' for endpoint adjudication, trials were generally scored a 2 in this domain. Of these

trials, only the RESPECT Trial specified the use of a validated symptom-detection questionnaires and automatic referral to mitigate referral bias, and therefore received a 1+.

Beyond these risks from a PROBE design, 3 trials had more serious concerns:

1. RESPECT had a substantial and differential drop out (albeit over a longer follow up time).

The dropout rate was 33.3% in the medical-therapy group and 20.8% in the PFO closure group, resulting in a significant between-group difference in the median duration of safety follow-up (2669 patient-years in the medical-therapy group vs. 3141 patient-years in the PFO closure group, p<.001). Higher risk patients appeared to drop out from the medical arm, potentially biasing toward the null.

2. The PC Trial had relatively high rates of drop out and also had some evidence of referral bias for endpoint adjudication.

Among 414 patients, 7 patients in the closure group and 11 in the medical-therapy group withdrew from the study; 24 and 31 others, respectively, were lost to follow-up.

There was a relatively low rate of referral for adjudication and differential rate of non-events (7 for medical therapy versus 2 for device) suggesting the possibility of less sensitive referral in the device arm.

3. The DEFENSE Trial did not have blinded outcome adjudication.

Small Study Effect

An assessment of small study effects by assessing funnel plot asymmetry. Trial sample sizes ranged from 120 (DEFENSE) to 980 (RESPECT). Visual inspection of the funnel plot for the six trials (where the CLOSE trial is treated as a single trial) did not suggest asymmetry. In addition, two formal tests for asymmetry were conducted. The test of asymmetry using the arcsin transformation for binary outcomes[24] was not statistically significant (p-value = 0.11). A similar linear regression test of asymmetry based on the log(hazard ratio) and standard error was also not significant (p-value = 0.59). These tests are generally

not recommended for meta-analyses with fewer than 10 studies and should be interpreted accordingly[25]. In two of the six trials included in our analysis there were no observed recurrent ischemic strokes in the device arm leading to unstable with-in trial estimated hazard ratios and standard errors. In an analysis excluding these trials (DEFENSE, CLOSE) the HR was 0.52 (95% CI, 0.35-0.78). These effect estimates reveal stability in our analysis of the primary outcome.