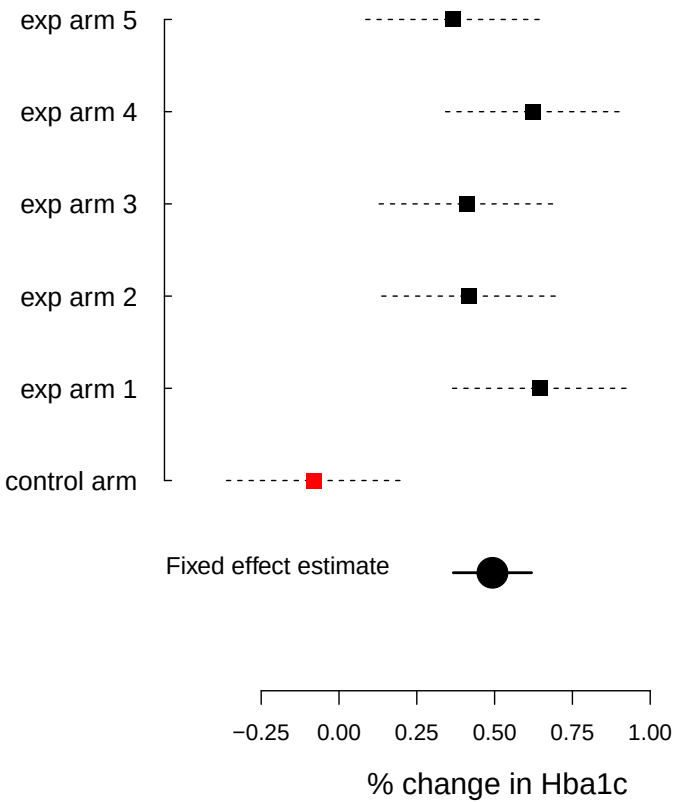
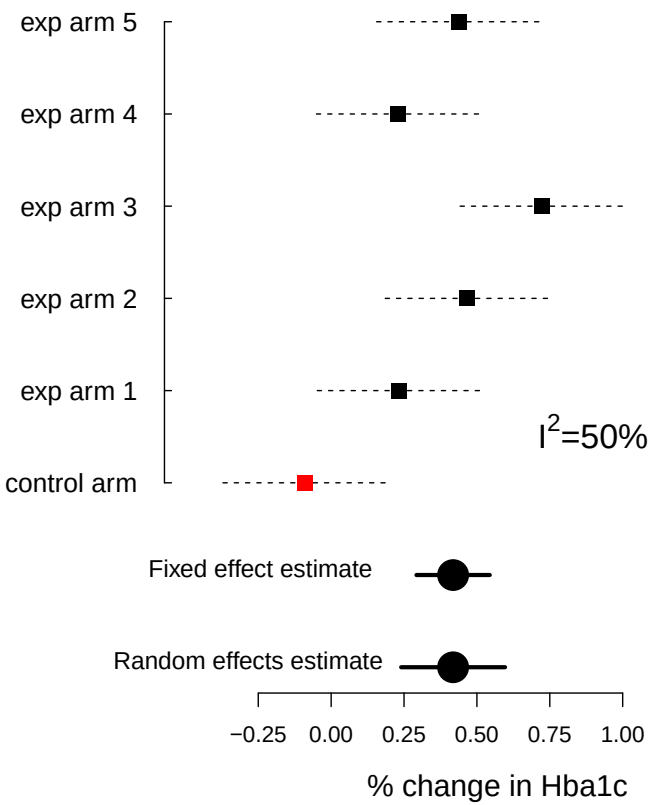
Illustration: δ=0.4%, σ=1.2% n=35 per-arm per-stage



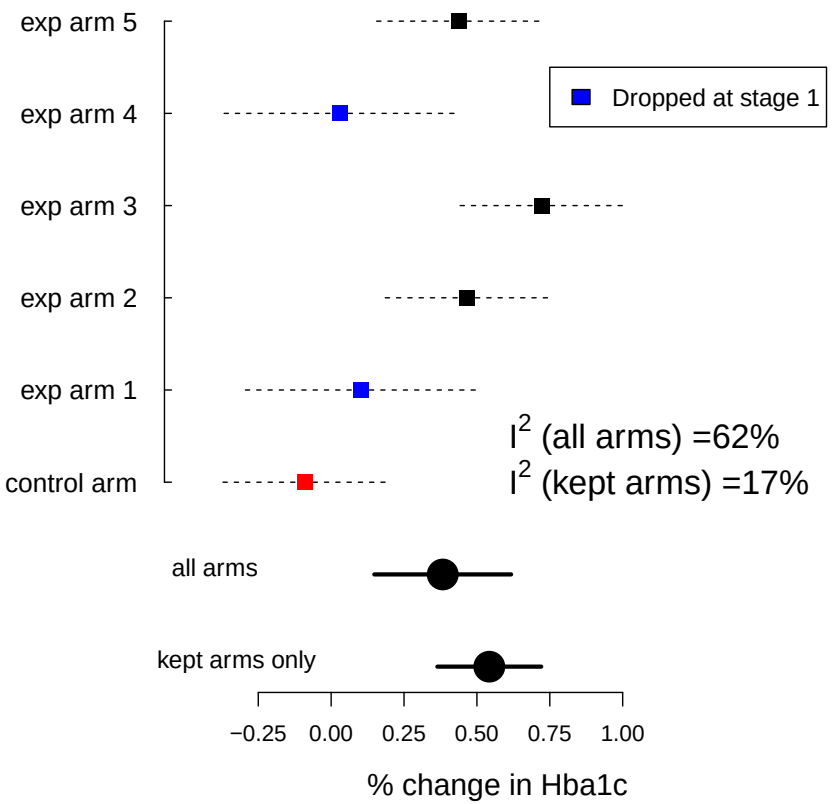
- Do all experimental treatments share a common effect?
- E.g experimental arm estimate $y_i = \mu + \epsilon_i$ where $\text{Var}(\epsilon_i)$ is variance of y_i
- Use Cochran's Q statistic to test null hypothesis.
- Here, p-value for Q is 0.5: No evidence to reject common effect (μ) hypothesis
- Pooled 'fixed effect' estimate for μ justified
- Compare pooled estimate's Confidence intervals to that of control group and declare class effective if no overlap

Heterogeneity between experimental treatments



- Perhaps treatments don't share a common effect?
- E.g experimental arm estimate $y_i = \mu_i + \epsilon_i$ where $\text{Var}(\mu_i)$ is between arm variation
- Between arm variation as a prop^n of total variation: $I^2 = 50\%$
- Q statistic p-value=0.09
- Pooled 'random-effects' estimate for μ (= mean of μ_is) arguably justified
- likely to be very similar to fixed effect estimate
- Wider confidence interval to acknowledge extra uncertainty. Lower power, but arguably right model

Issues surrounding dropped treatments



- Assume arm 1 & 4 dropped at interim (after 35 patients)
- Dropped trials have less precise estimates than kept trials
- Should we exclude dropped arms when estimating pooled effect?
- **Exclude**: PROS: Remaining arm results more homogeneous. Likely to opt for a fixed effect estimate. CONS: Remaining arms potentially biased, throwing away information
- **Include**: PROS: Using all available information. CONS: Confidence interval may still be wider due to use of random effects model.