Appendix

Kernel density methodology aims to estimate the density function of a random variable **θ** from a random sample θ_i without assuming that the function belongs to a known parametric family. By constructing a non-parametric density of treatment effects, we can gain insight into the natural variability of treatment effects across studies.

Given n number of studies and individual study treatment effects $\theta_1,\ldots,\theta_n$ a weighted kernel density estimator of treatment effects is defined as

$$\hat{f}(\theta) = \sum_{i=1}^{n} w(\theta_i) K_h(\theta - \theta_i) \quad (0.1)$$

where $K_h(.)$ is a kernel function and h is the bandwidth which controls the smoothness of the density estimate. Essentially, kernel density estimate (KDE) is a continuous histogram whose blocks are centralized in each of the data points from where the density is estimated. The kernel function defines the shapes of the peaks of the observed data so that the estimator is the sum of the peaks.

Properties of the kernel function K(u) partially determine the properties of KDE, such as differentiability and continuity, so $K_h(.)$ is usually chosen to be a symmetric density function; $w(.)$ is a re-weighting function which is used to control the roles of different θ_i. In the context of this, $w(.)$ was formulated to incorporate study-level sampling errors σ_i and between-study heterogeneity τ^2. This is equivalent to weighting in standard random effects meta-analysis.

The most important issue in KDE is the bandwidth selection. There have been several approaches to finding the “optimal” bandwidth. We considered the approach based on minimizing the mean integrated square error, under which the approximate optimal bandwidth is given by

$$h = 0.9\ \min(s_w, IQR_w/1.34)\ n^{-1/5}, \quad (0.2)$$

for s_w = sample standard deviation and IQR_w = sample inter-quartile range. A drawback of using a fixed bandwidth given in (0.2) is that where the data are dense will be masked or spurious noise will appear where the data are sparse. To account for data clustering in estimating the bandwidth, the adaptive kernel density estimates can be calculated using the following three step procedure [13]:

Step 1: Find a pilot density estimate $\hat{f}$ by (0.1) with bandwidth h from (0.2).

Step 2: Define local bandwidth factor $L_i = (\frac{g}{\hat{f}})^{\alpha}$, where $\log g = n^{-1} \sum log \hat{f}$.

Step 3: Define the adaptive kernel estimate $\tilde{f}$ by

$$\tilde{f} = \sum_{i=1}^{n} \frac{w(\theta_i)}{hL_i} K_h(\frac{\theta - \theta_i}{hL_i}) \quad (0.3)$$

Literature has shown that the pilot estimate in Step 1 is not crucial.