In S1 Fig the rating results as reported by Sonuga-Barke et al. [1] were included, except for Schmidt et al.’s teacher ratings results, which were replaced by the test observation rating results [2]. Since Sonuga-Barke et al. supplied graphical results only (see Figure 3.A, page 283 of Sonuga-Barke et al.’s paper [1]) we used MS Publisher’s ruler to obtain the effect size and lower and upper level of the 95% CI of the blinded measurement data. For three reasons we replaced the teacher measurements provided in the fifth study [2] and included by Sonuga-Barke et al. [1], by the test observation measurements also provided by Schmidt et al. [2]:

First, the teacher ratings provided by Schmidt et al. [2] appeared to be biased. These ratings were generated in a specialist setting and could not even detect behavioral problems at baseline (see page 92, Table 2, of Schmidt et al.’s paper [2]). Probably, as has been suggested by Sonuga-Barke et al. [7], specialized teacher settings might be therapeutically beneficial already. In addition, ‘classroom ratings were not even able to demonstrate the well-established treatment effects of stimulant therapy’ [2]. For good reason Schmidt et al. excluded these ratings from further analyses [2].

Second, according to the guidelines described by Sonuga-Barke et al. [1] observer ratings should be considered ‘better probably blinded’ than teacher ratings, and thus, if both ratings were available, the observer ratings should be included (see also Results, section ‘Sub-analysis based on the FFD meta-analysis by Sonuga-Barke et al.’, bullet point 2).

Third, reproducing the meta-analytic calculations of Sonuga-Barke et al. [1] including Schmidt et al.’s test observation measurements rather than the specialized teacher ratings [2] may reduce bias in the meta-analytic estimates.

The calculations provided in S1 Fig resulted in an effect size of 0.57 (95% CI: 0.07–1.06, p = 0.024, I² = 67%), which is commensurate with the results of Sonuga-Barke et al. [1] (ES = 0.51, 95% CI: -0.02–1.04, p = 0.06, I² = 72%). However, if an intervention’s efficacy is inferred from p-values being < 0.05,
then the inclusion of either Schmidt et al.’s teacher ratings ($p = 0.06$) or test observation ratings ($p = 0.024$) would potentially generate different conclusions. The following three statements in the abstract of Sonuga-Barke et al.’s paper [1] suggest that statistical effects are considered more important than clinical effects:

1) The significant effects of most treatments attenuated to non-significant levels when evaluating the blinded ratings only,

2) PUFA supplementation ($ES = 0.16$, $p = 0.04$) ‘produced small but significant reductions in ADHD symptoms even with probably blinded assessments, although the clinical significance of these effects remains to be determined’,

3) ‘Better evidence for efficacy from blinded assessments is required’ for a FFD ($ES = 0.51$, $p = 0.06$) before this intervention ‘can be supported as treatments for core ADHD symptoms’.

These statements convey that, although effect sizes are reported as well, the conclusions are p-value based. Comparison of the results reported in S1 Fig and in Figure 3.A. provided by Sonuga-Barke et al. [1], illustrates the limitation of relying on p-values, a topic which is further addressed in the discussion section (see ‘Differences in conclusions between our and previous reviews’).

References


