14.6. Training children to avoid sexual abuse

This example is a re-analysis of a systematic review published by Walsh, Zwi, Woolfenden, and Shlonsky (2015). References in the text refer to Figure 103 through Figure 106.

Overview

The analysis is based on eighteen studies that evaluated the effect of classroom teaching on students' ability to identify and respond to potential sexual abuse. Each study compared knowledge, as measured by a test, for students who had been trained vs. those in a control condition. The effect size is the standardized difference in means, Hedges' *g*. The results of this analysis will be generalized to comparable studies, and so the random-effects model was employed for the analysis.

Does training increase knowledge?

The mean effect size (Hedges' *g*) is 0.616 [A]. On average, students who received the training scored 0.616 standard deviations higher than those who had not. On a scale with a standard deviation of 10 points, this would correspond to a difference of 6.16 points. The confidence interval for the effect size is 0.450 to 0.782 [B], which tells us that the standardized mean difference in the universe of comparable studies could fall anywhere in this range. This range does not include an effect size of zero, which tells us that the true effect size is probably not zero. Similarly, the *Z*-value for testing the null hypothesis (that the mean effect size is 0.0) is 7.281, with a corresponding *p*-value of < 0.001 [C]. Using the Knapp-Hartung adjustment, $t = 7.05$, $df = 17$, $p < 0.001$, and the 95% confidence interval is 0.432 to 0.801. We can reject the null hypothesis, and conclude that (on average) the training does increase knowledge in the universe of populations which are comparable to those in the analysis. Given the dispersion in effects (as discussed below), it is important to recognize that the mean effect size applies to this particular mix of studies, and would be different for another mix of populations, durations of the intervention, and so on.

How much does the effect size vary across studies?

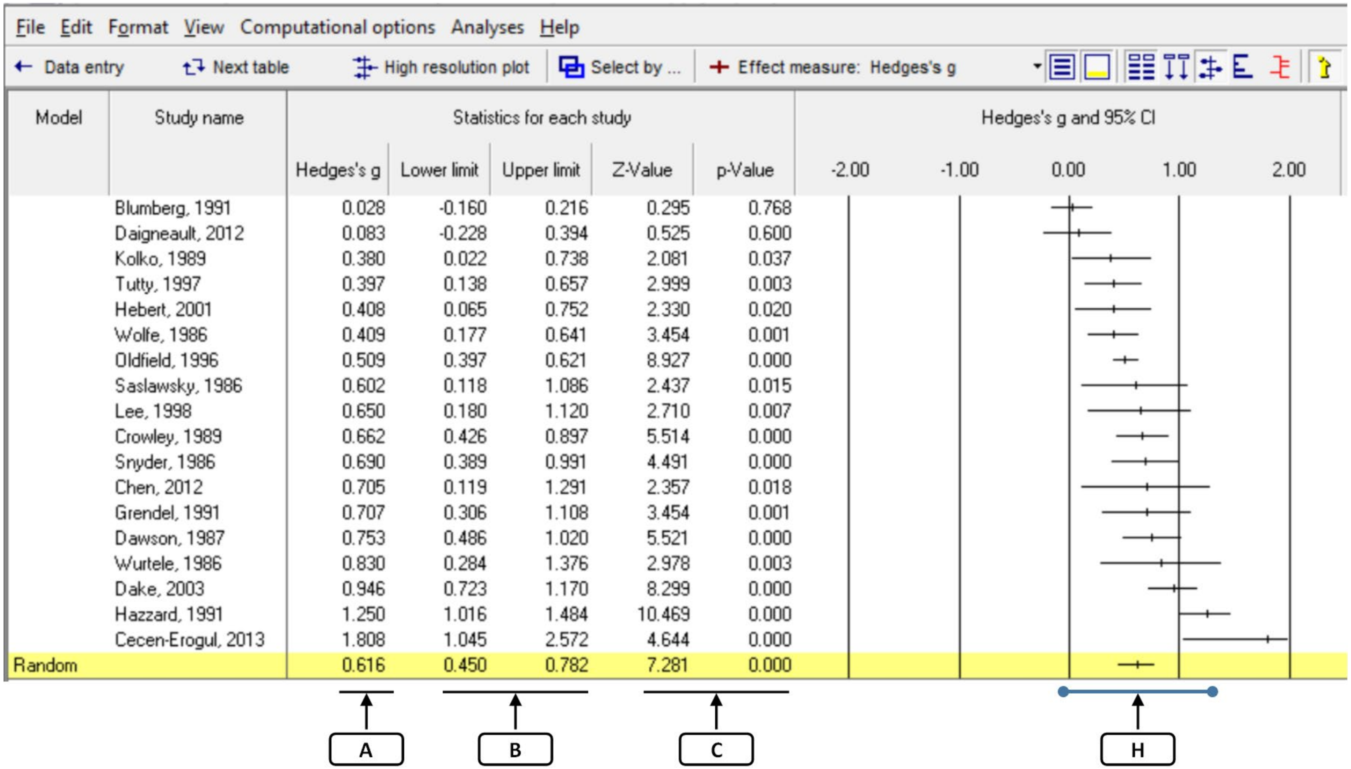
The *Q*-statistic provides a test of the null hypothesis that all studies in the analysis share a common effect size. If all studies shared the same effect size, the expected value of Q would be equal to the degrees of freedom (the number of studies minus 1). The Q-value is 105.930 with 17 degrees of freedom and a p-value of < 0.001. We reject the null hypothesis that the true effect size is identical in all the studies [D]. The I^2 statistic is 84%, which tells us that 84% of the variance in observed effects reflects variance in true effects rather than sampling error [E]. The variance of true effects (T^2) is 0.097 [F], and the standard deviation of true effects (T) is 0.312 [G]. The 95% prediction interval is −0.069 to +1.301 [H]. In the universe of populations represented by these studies, the true effect size in 95% of cases will fall somewhere in this range. The mean effect size is moderately impressive, so on average the extent of knowledge is improved by an amount that may well have a substantive impact. However, the dispersion of effects about this mean is substantial. There are some populations where the impact is very strong, some where it is moderate, and some where it is trivial.

File Edit Format View Insert Identify Tools Computational options Analyses Help

Run analyses

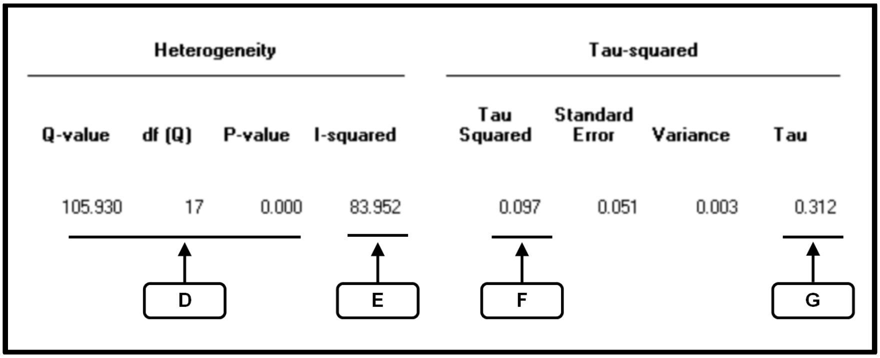
	Study name	Trained Mean	Trained Std-Dev	Trained Sample size	Control Mean	Control Std-Dev	Control Sample size	Effect direction	Hedges's g	Std Err	Variance
1	Blumberg, 1991	18.250	3.670	322	18.150	3.260	164	Auto	0.028	0.096	0.009
2	Daigneault, 2012	13.700	5.020	70	13.300	4.590	90	Auto	0.083	0.159	0.025
3	Kolko, 1989	12.400	2.400	213	11.500	2.100	35	Auto	0.380	0.183	0.033
4	Tutty, 1997	8.500	0.900	117	8.100	1.100	114	Auto	0.397	0.132	0.018
5	Hebert, 2001	8.540	2.020	59	7.680	2.150	74	Auto	0.408	0.175	0.031
6	Wolfe, 1986	5.300	1.320	145	4.710	1.550	145	Auto	0.409	0.118	0.014
7	Oldfield, 1996	26.690	4.950	658	24.080	5.300	611	Auto	0.509	0.057	0.003
8	Saslawsky, 1986	11.240	2.380	33	9.790	2.380	34	Auto	0.602	0.247	0.061
9	Lee, 1998	8.970	1.820	38	7.790	1.770	34	Auto	0.650	0.240	0.057
10	Crowley, 1989	24.520	1.700	157	22.760	3.440	136	Auto	0.662	0.120	0.014
11	Snyder, 1986	26.270	5.000	89	23.290	3.460	89	Auto	0.690	0.154	0.024
12	Chen, 2012	4.350	1.071	23	3.520	1.238	23	Auto	0.705	0.299	0.089
13	Grendel, 1991	10.670	1.050	51	9.830	1.300	49	Auto	0.707	0.205	0.042
14	Dawson, 1987	78.750	18.710	96	64.680	18.577	141	Auto	0.753	0.136	0.019
15	Wurtele, 1986	11.530	1.920	53	9.720	2.760	18	Auto	0.830	0.279	0.078
16	Dake, 2003	12.300	2.000	166	10.200	2.400	175	Auto	0.946	0.114	0.013
17	Hazzard, 1991	20.600	3.670	286	15.400	5.180	113	Auto	1.250	0.119	0.014
18	Cecen-Erogul, 2013	8.660	1.280	18	6.160	1.420	18	Auto	1.808	0.389	0.152
19											

Figure 103 | Impact of training on preventing sexual abuse | Data entry in CMA



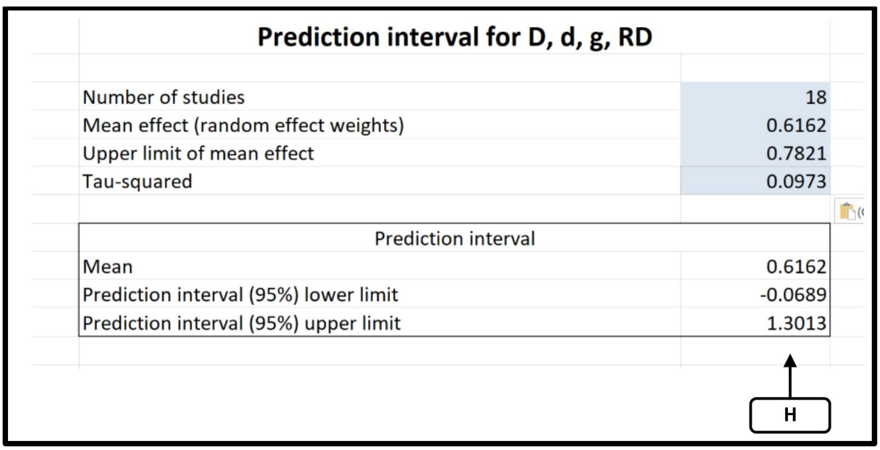
Model	Study name	Hedges's g	Lower limit	Upper limit	Z-Value	p-Value
	Blumberg, 1991	0.028	-0.160	0.216	0.295	0.768
	Daigneault, 2012	0.083	-0.228	0.394	0.525	0.600
	Kolko, 1989	0.380	0.022	0.738	2.081	0.037
	Tutty, 1997	0.397	0.138	0.657	2.999	0.003
	Hebert, 2001	0.408	0.065	0.752	2.330	0.020
	Wolfe, 1986	0.409	0.177	0.641	3.454	0.001
	Oldfield, 1996	0.509	0.397	0.621	8.927	0.000
	Saslawsky, 1986	0.602	0.118	1.086	2.437	0.015
	Lee, 1998	0.650	0.180	1.120	2.710	0.007
	Crowley, 1989	0.662	0.426	0.897	5.514	0.000
	Snyder, 1986	0.690	0.389	0.991	4.491	0.000
	Chen, 2012	0.705	0.119	1.291	2.357	0.018
	Grendel, 1991	0.707	0.306	1.108	3.454	0.001
	Dawson, 1987	0.753	0.486	1.020	5.521	0.000
	Wurtele, 1986	0.830	0.284	1.376	2.978	0.003
	Dake, 2003	0.946	0.723	1.170	8.299	0.000
	Hazzard, 1991	1.250	1.016	1.484	10.469	0.000
	Cecen-Erogul, 2013	1.808	1.045	2.572	4.644	0.000
Random		0.616	0.450	0.782	7.281	0.000

Figure 104 | Preventing sexual abuse | Standardized mean difference > 0 favors training



Heterogeneity				Tau-squared			
Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
105.930	17	0.000	83.952	0.097	0.051	0.003	0.312

Figure 105 | Preventing sexual abuse | Heterogeneity statistics in CMA



Prediction interval for D, d, g, RD

Number of studies	18
Mean effect (random effect weights)	0.6162
Upper limit of mean effect	0.7821
Tau-squared	0.0973

Prediction interval	
Mean	0.6162
Prediction interval (95%) lower limit	-0.0689
Prediction interval (95%) upper limit	1.3013

Figure 106 | Preventing sexual abuse | Prediction interval in Excel