Supplementary Material

Definitions of coding categories for estimation strategies

Descriptions of estimation strategies were classified according to a pre-determined coding scheme developed on the basis of pilot data and supplemented with additional categories developed inductively using a subsample of strategy descriptions from the main study (Table S1)[13]. A second coder double-coded one third of the strategies for reliability (Cohen's κ=0.73). Estimation strategies are summarised descriptively.

Table S1: Definitions of coding strategies for estimates of treatment and placebo effects.

Estimation Strategy	Definition
When one number shown (e.g., point estimate condition)	
Numbers as shown	Provide the exact number shown.
When two numbers shown (e.g., imprecision or conflicting estimates)	
Upper number shown	Provide the exact upper number shown.
Lower number shown	Provide the exact lower number shown
Median estimate	Provide the median (average) of the two numbers shown.
Any case	
Anchoring	Make a small adjustment – within +/- 5 range – to the a) lower or b) upper of two values, or c) given or median of two values.
10+ below/above lowest/ highest number	Provide an estimate that is 10+ points above or below the lowest or highest number shown (or placebo for the *lack of evidence*).
Half of placebo	Provide an estimate that is half the placebo value.
Placebo for treatment	Used placebo as treatment estimate.
Guessed or not codeable	Strategy did not fit any other coding category or was unclear
*Additional strategies**	
Average placebo and treatment	Averaged the placebo and treatment value
Average all studies	Averaged two treatment and placebo values
50 percent	Stated that they gave a 50/50 estimate
Anchor on placebo	Make a small adjustment – within +/-5 range – to the placebo value
Strategy did not match	The stated strategy did not match the number provided

Note: A category *Half lowest number shown* was removed in place of the categories 10+ below/above lowest/highest estimate.; *Strategies were added to the coding scheme after sub-sample of coding was completed.

Distributional perceptions.

At the end of the study, participants in the *imprecision* and *conflicting estimates* condition completed a multiple choice question on whether they perceived the numbers in the range [two different study estimates] to be equally likely, or values in the middle of the range [middle of the two estimates], at the low end of the range [closer to the lower estimate], or high end of the range [closer to the higher estimate] to be more likely. These perceptions relate to uniform, normal, positively or negatively skewed distributions, respectively. The item was adapted from Dieckmann et al. (2015).

---

[13]As it was expected that not all participants would provide clear descriptions that could be easily coded (e.g., "*I took it from the graph*"), numerical estimates were also classified according to the coding categories. Results were largely consistent, except for a higher proportion of "not codeable" responses in the *lack of evidence* condition

The majority of participants in the *imprecision* and *conflicting estimates* conditions interpreted the numbers they received as consistent with a normal (63.2% and 62.3%, respectively) or a uniform distribution (20.1% and 23.9%). The remaining participants perceived values closer to the lower estimate (10.5% and 10.1%) as slightly more likely than values closer to the higher estimate (5.7% and 3.0%), consistent with a positively and negatively skewed distribution, respectively. Responses were largely consistent across display formats (see S1).
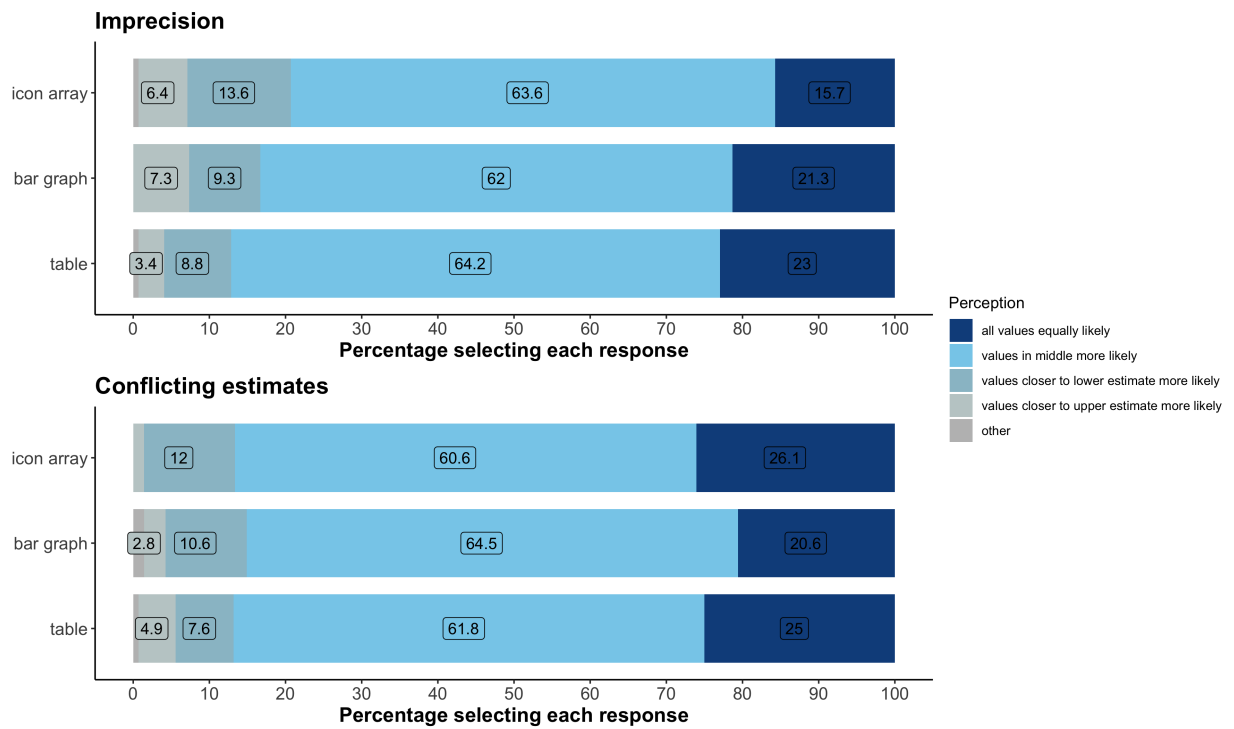

Figure S1: Distribution perceptions for *imprecision* and *conflicting estimates* for each display format. Numbers in bars refer to percentages. Percentages for "other" responses were small and are not shown in numbers.

Estimated Marginal Means: Uncertainty Type

Table S2: Estimated marginal mean differences for uncertainty type for each outcome measure.

	contrast	difference estimate	SE	p.value	lower 95%CI	upper 95%CI	effect size
	Variation						
1	point estimate - imprecision	-0.95	0.49	0.20	-2.20	0.29	-0.17
2	point estimate - conflicting	-0.83	0.49	0.32	-2.09	0.42	-0.15
3	point estimate - lack of evidence	-12.40	0.49	0.00	-13.66	-11.14	-2.20
4	imprecision - conflicting	0.12	0.49	0.99	-1.13	1.37	0.02
5	imprecision - lack of evidence	-11.45	0.49	0.00	-12.70	-10.19	-2.03
6	conflicting - lack of evidence	-11.57	0.49	0.00	-12.83	-10.30	-2.06
	Subjective evaluations						
1	point estimate - imprecision	0.00	0.05	1.00	-0.12	0.13	0.01
2	point estimate - conflicting	0.06	0.05	0.57	-0.06	0.19	0.16
3	point estimate - lack of evidence	0.54	0.05	0.00	0.41	0.66	1.38
4	imprecision - conflicting	0.06	0.05	0.61	-0.06	0.18	0.15
5	imprecision - lack of evidence	0.54	0.05	0.00	0.41	0.66	1.37
6	conflicting - lack of evidence	0.48	0.05	0.00	0.35	0.60	1.22
	Certainty perceptions						
1	point estimate - imprecision	-0.02	0.06	0.99	-0.17	0.14	-0.04
2	point estimate - conflicting	0.10	0.06	0.32	-0.05	0.26	0.26
3	point estimate - lack of evidence	0.40	0.06	0.00	0.24	0.55	1.01
4	imprecision - conflicting	0.12	0.06	0.19	-0.03	0.27	0.30
5	imprecision - lack of evidence	0.41	0.06	0.00	0.26	0.57	1.05
6	conflicting - lack of evidence	0.30	0.06	0.00	0.14	0.45	0.75
	Trustworthiness						
1	point estimate - imprecision	-0.01	0.06	1.00	-0.15	0.14	-0.01
2	point estimate - conflicting	0.13	0.06	0.09	-0.01	0.28	0.29
3	point estimate - lack of evidence	0.29	0.06	0.00	0.15	0.44	0.65
4	imprecision - conflicting	0.14	0.06	0.07	-0.01	0.28	0.31
5	imprecision - lack of evidence	0.30	0.06	0.00	0.16	0.45	0.67
6	conflicting - lack of evidence	0.16	0.06	0.02	0.02	0.31	0.36
	Behavioral intentions						
1	point estimate - imprecision	0.26	0.16	0.37	-0.15	0.67	0.13
2	point estimate - conflicting	0.34	0.16	0.16	-0.08	0.76	0.17
3	point estimate - lack of evidence	-0.12	0.16	0.88	-0.54	0.30	-0.06
4	imprecision - conflicting	0.08	0.16	0.96	-0.34	0.49	0.04
5	imprecision - lack of evidence	-0.38	0.16	0.09	-0.80	0.04	-0.19
6	conflicting - lack of evidence	-0.46	0.16	0.03	-0.88	-0.04	-0.23
		Odds Ratio	SE	p.value	lower 95%CI	upper 95%CI	
	Recall						
1	point estimate / imprecision	1.41	0.39	0.60	0.69	2.87	
2	point estimate / conflicting	1.69	0.47	0.22	0.84	3.43	
3	point estimate / lack of evidence	3.18	0.87	0.00	1.58	6.41	
4	imprecision / conflicting	1.20	0.32	0.90	0.60	2.39	
5	imprecision / lack of evidence	2.25	0.60	0.01	1.14	4.46	
6	conflicting / lack of evidence	1.88	0.49	0.08	0.96	3.68	

.

Results are averaged over the levels of: display format. Degrees-of-freedom method: asymptotic. P value adjustment: Tukey method for comparing a family of 4 estimates. Effect size = Cohen's *d*.

Estimated Marginal Means: Display Format

Table S3: Estimated marginal mean differences for display format for each outcome measure.

	contrast	difference estimate	SE	p.value	lower 95%CI	upper 95%CI	effect size
	Variation						
1	table - bar graph	0.08	0.42	0.98	-0.91	1.07	0.01
2	table - icon array	0.12	0.42	0.95	-0.86	1.11	0.02
3	bar graph - icon array	0.05	0.42	0.99	-0.95	1.04	0.01
	Subjective evaluations						
1	table - bar graph	0.04	0.04	0.59	-0.06	0.14	0.10
2	table - icon array	-0.01	0.04	0.98	-0.10	0.09	-0.02
3	bar graph - icon array	-0.05	0.04	0.49	-0.15	0.05	-0.12
	Certainty perceptions						
1	table - bar graph	0.07	0.05	0.40	-0.05	0.19	0.17
2	table - icon array	0.01	0.05	0.98	-0.11	0.13	0.03
3	bar graph - icon array	-0.06	0.05	0.54	-0.18	0.07	-0.14
	Trustworthiness						
1	table - bar graph	0.07	0.05	0.31	-0.04	0.18	0.16
2	table - icon array	0.02	0.05	0.94	-0.10	0.13	0.04
3	bar graph - icon array	-0.05	0.05	0.50	-0.17	0.06	-0.12
	Behavioral intentions						
1	table - bar graph	0.22	0.14	0.26	-0.11	0.55	0.11
2	table - icon array	0.01	0.14	1.00	-0.32	0.33	0.00
3	bar graph - icon array	-0.21	0.14	0.29	-0.54	0.12	-0.11
		Odds Ratio	SE	p.value	lower 95%CI	upper 95%CI	
	Recall						
1	table / bar graph	0.91	0.21	0.92	0.53	1.58	
2	table / icon array	1.42	0.33	0.28	0.83	2.44	
3	bar graph / icon array	1.55	0.36	0.14	0.90	2.69	

.
Results are averaged over the levels of: uncertainty type. Degrees-of-freedom method: asymptotic. P value adjustment: Tukey method for comparing a family of 4 estimates. Effect size = Cohen's *d*.

Secondary Analyses: Linear mixed effects models excluding *lack of evidence*

Table S4: Results of linear mixed effects models on variation, subjective evaluations, certainty perceptions, trust, and behavioral intentions. (excluding *lack of evidence* condition)

	Variation[a] Estimate [95%CI]	Subjective evaluations[b] Estimate [95%CI]	Certainty perceptions[b] Estimate [95%CI]	Trust[b] Estimate [95%CI]	Behavioral intentions[c] Estimate [95%CI]
Fixed effects					
Intercept	9.17 [7.96; 10.38]	4.25 [4.14; 4.37]	4.13 [3.98; 4.28]	4.07 [3.92; 4.22]	8.72 [8.30; 9.12]
Uncertainty type					
Point estimate	referent	referent	referent	referent	referent
Imprecision	0.97 [0.07; 1.85]	0.00 [−0.09; 0.09]	0.02 [−0.09; 0.14]	0.01 [−0.09; 0.13]	−0.25 [−0.55; 0.06]
Conflicting estimates	0.84 [−0.03; 1.71]	−0.05 [−0.15; 0.04]	−0.10 [−0.21; 0.01]	−0.13 [−0.24; −0.02]	−0.33 [−0.62; −0.01]
Display format					
Table	referent	referent	referent	referent	referent
Bar graph	−0.04 [−0.91; 0.78]	−0.06 [−0.14; 0.03]	−0.09 [−0.21; 0.02]	−0.11 [−0.22; 0.00]	−0.27 [−0.59; 0.03]
Icon array	0.25 [−0.65; 1.07]	−0.03 [−0.12; 0.06]	−0.03 [−0.14; 0.08]	−0.05 [−0.15; 0.07]	0.06 [−0.25; 0.38]
Covariates					
Graph literacy[d]	−1.99 [−2.31; −1.66]	0.02 [−0.01; 0.05]	−0.13 [−0.17; −0.08]	−0.08 [−0.12; −0.04]	−0.20 [−0.31; −0.08]
Numeracy[e]	−0.59 [−0.92; −0.28]	0.02 [−0.01; 0.06]	−0.01 [−0.06; 0.03]	0.04 [0.00; 0.08]	0.03 [−0.08; 0.15]
Treatment effect size					
Moderate effect	referent	referent	referent	referent	referent
Small effect	−0.60 [−0.98; −0.25]	−0.08 [−0.11; −0.05]	−0.10 [−0.13; −0.07]	−0.05 [−0.08; −0.02]	−1.67 [−1.81; −1.53]
Random effects					
Intercept σ^2	30.65	0.37	0.62	0.56	3.77
Residual	23.40	0.13	0.15	0.19	3.25

Note: referent = reference group; Estimates are unstandardized coefficients. Square brackets indicate bootstrapped confidence intervals. [a]Mean absolute deviations from each group's median estimate; [b]5-point Likert scales ranging from 'not at all'–'very'. [c]11-point Likert scale 'I would definitely not – definitely take [medication]. [d]Scores range 0-4; higher scores indicate greater graph literacy. [e]Scores range 1-4; higher scores indicate greater numeracy.

Recall analyses

Table S5: Results of general linear mixed effects models for recall, including the *lack of evidence* condition (left) and excluding *lack of evidence* condition (right)

	Recall[a]: including *lack of evidence* Odds Ratio [95%CI]	Recall[a]: excluding *lack of evidence** Odds Ratio [95%CI]
Fixed effects		
Intercept	5.32	4.16
	[6.00; 59.18]	[8.11; 309.40]
Uncertainty type		
Point estimate	referent	referent
Imprecision	0.71	0.70
	[0.42; 1.25]	[0.41; 1.26]
Conflicting estimates	0.59	0.58
	[0.34; 1.09]	[0.36; 1.07]
Lack of evidence	0.31	-
	[0.17; 0.54]	
Display format		
Table	referent	referent
Bar graph	1.09	1.30
	[0.64; 1.69]	[0.72; 2.12]
Icon array	0.70	0.91
	[0.42; 1.12]	[0.53; 1.56]
Covariates		
Graph literacy[b]	3.30	3.36
	[2.80; 4.28]	[2.12; 3.87]
Numeracy[c]	1.29	1.35
	[1.03; 1.49]	[1.03; 1.54]
Recall value		
Recall placebo	referent	referent
Recall treatment	0.43	0.61
	[0.30; 0.49]	[0.40; 0.73]
Random effects		
Intercept σ^2	8.19	9.72

Note: referent = reference group; Estimates are odds ratios and square brackets indicate bootstrapped confidence intervals. [a]Recall of treatment and placebo numbers for both medications were coded such that each participant received a binary score of correct (vs. incorrect) for each of the four numbers (two medications each with a treatment value and placebo value). [b]Scores range 0-4 with higher scores indicating greater graph literacy. [c]Scores range from 1-4 with higher scores indicating greater numeracy. *A model including treatment effect did not improve model fit ($\chi^2(1)$=1.96, p=.161).