eLIFE
elifesciences.org

Figures and figure supplements

A meta-analysis of threats to valid clinical inference in preclinical research of sunitinib

Valerie C Henderson, et al.

A

Internal Validity Characteristics	Experiments (n=158)
Exact sample size given for all groups	120 (76%)
Performance of sample size calculation	0 (0%)
Randomly allocated to treatment	58 (37%)
Concealed allocation	0 (0%)
Blinded outcome assessment	0 (0%)
Used named inferential statistical test	117 (74%)
Addressed animal flow through experiment	21 (13%)
Evaluated dose-response (≥3 doses)	9 (6%)
Construct Validity Characteristics	
Species	
Mouse	156 (99%)
Rat	2 (1%)
Age	
Pediatric/Juvenile	78 (49%)
Adult	25 (16%)
Aged	0 (0%)
Immune Status	
Immunocompetent	30 (19%)
Immunocompromised	128 (81%)
Sex	
Male	12 (8%)
Female	104 (66%)
Model Type	
Human xenograft	110 (70%)
Allografts	42 (27%)
Genetically engineered	4 (2%)
Other	2 (1%)
Experiment type	
Late-Stage Tumour Model	63 (40%)
Early-Stage Tumour Model	91 (58%)
Evidence for causal mechanism	
Molecular	66 (42%)
Physiological	123 (78%)
Functional/Clinical/Behavioural	0 (0%)

B

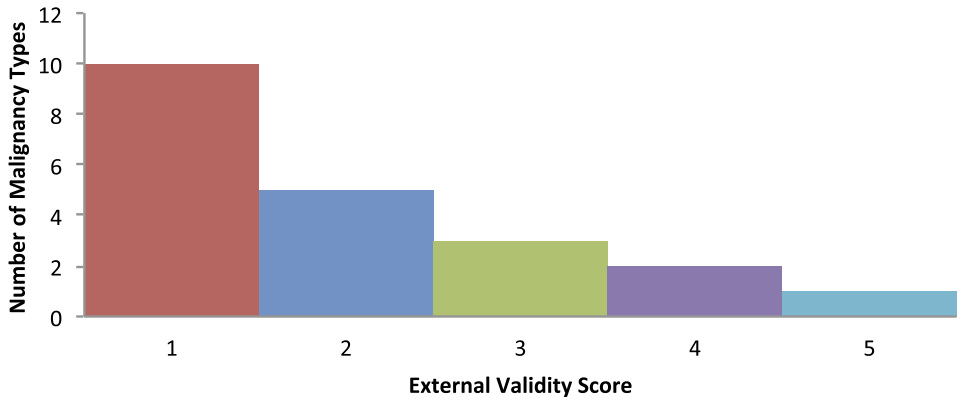

Figure 1. Descriptive analysis of (**A**) internal, construct, and (**B**) external validity design elements. External validity scores were calculated for each malignancy type tested, according to the formula: number species used + number of models used; an extra point was assigned if a malignancy type tested more than one species *and* more than one model.

DOI: 10.7554/eLife.08351.003

A

Internal Validity Characteristics		Included Experiments (n=332)
Exact sample size given for all groups		249 (75%)
Performance of sample size calculation		0 (0%)
Randomly allocated to treatment		144 (43%)
Concealed allocation		50 (15%)
Blinded outcome assessment		51 (15%)
Uses named inferential statistical test		264 (80%)
Adressed animal flow through experiment		79 (24%)
Evaluated dose-response (>3 doses)		23 (7%)
Construct Validity Characteristics		
Species	Mouse	325 (98%)
	Rat	3 (1%)
	Zebrafish	4 (1%)
Age	Pediatric/Juvenile	147 (44%)
	Adult	57 (17%)
	Aged	0 (0%)
Immune Status	Immunocompetent	73 (22%)
	Immunocompromised	259 (78%)
Sex	Male	28 (8%)
	Female	211 (64%)
	Mixed	3 (1%)
Model type	Human xenograft	237 (71%)
	Allografts (syngeneic, allogeneic)	65 (20%)
	Genetically engineered	24 (7%)
	Other	6 (2%)
Experiment type	Late-Stage Tumour Growth Model	150 (45%)
	Early-Stage Tumour Growth Model	134 (40%)
	Survival Analysis	39 (12%)
Evidence for causal mechanism	Molecular	114 (34%)
	Physiological	219 (66%)
	Functional/Clinical/Behavioural	1 (0.3%)

B

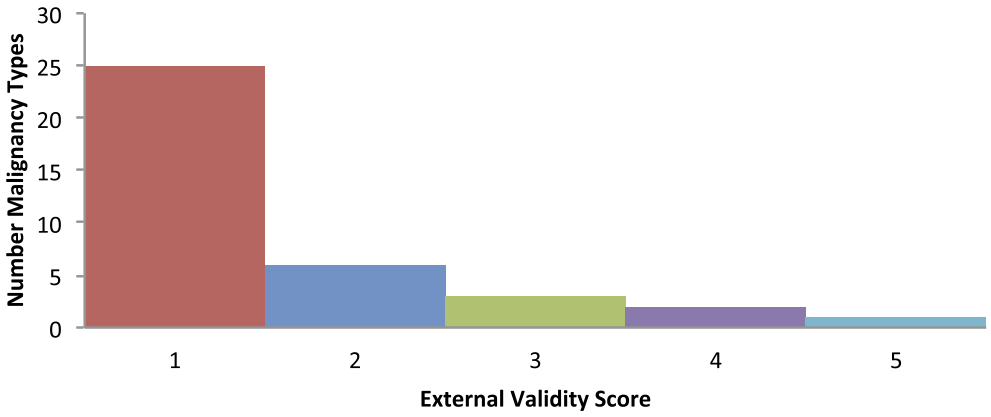

Figure 1—figure supplement 1. Descriptive analysis of (**A**) internal, construct, and (**B**) external validity design elements for all experiments (n = 332) extracted for validity data parameters.
DOI: 10.7554/eLife.08351.005

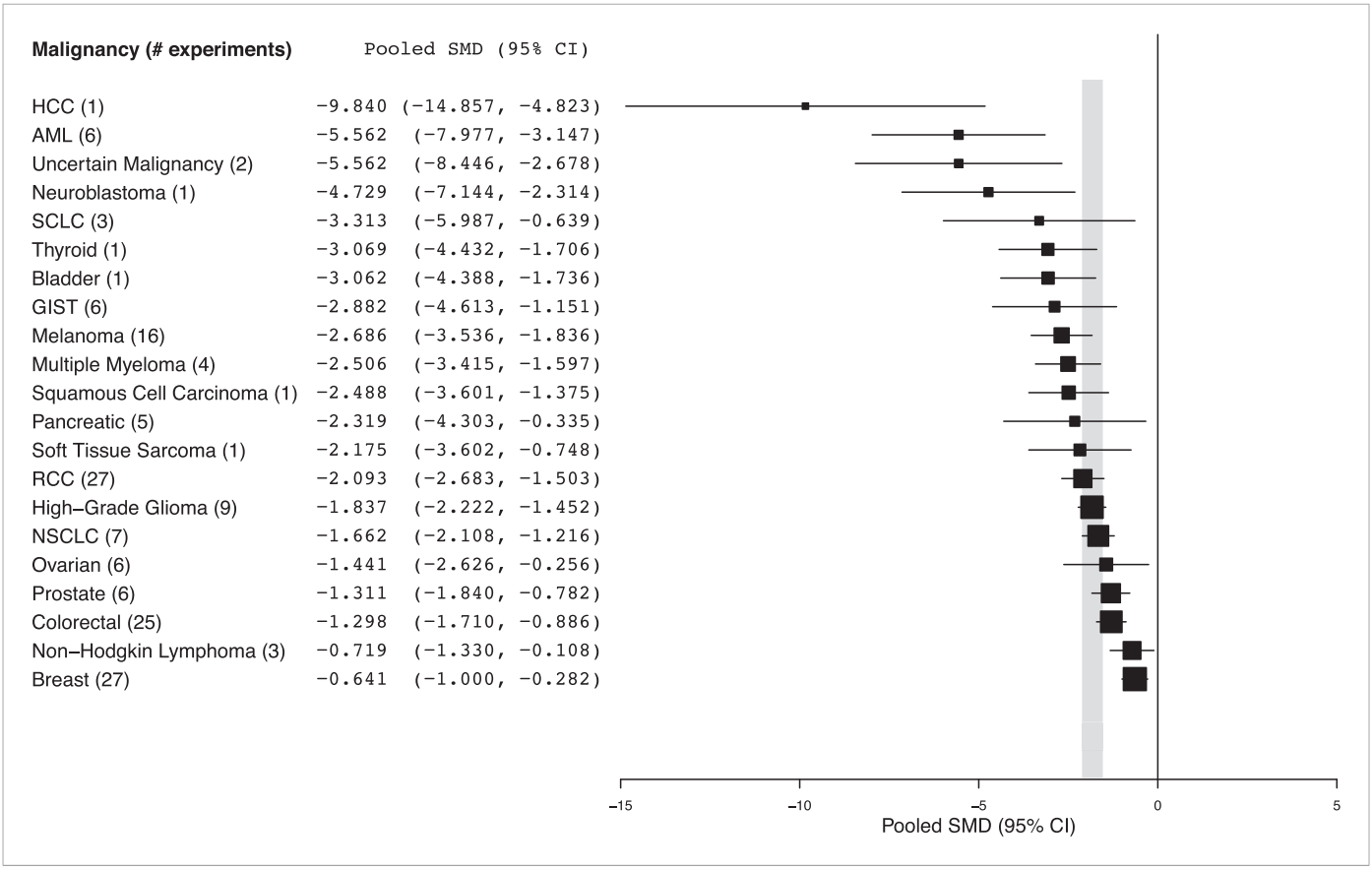

Figure 2. Summary of pooled SMDs for each malignancy type. Shaded region denotes the pooled standardized mean difference (SMD) and 95% confidence interval (CI) (−1.8 [−2.1, −1.6]) for all experiments combined at the last common time point (LCT).

DOI: 10.7554/eLife.08351.008

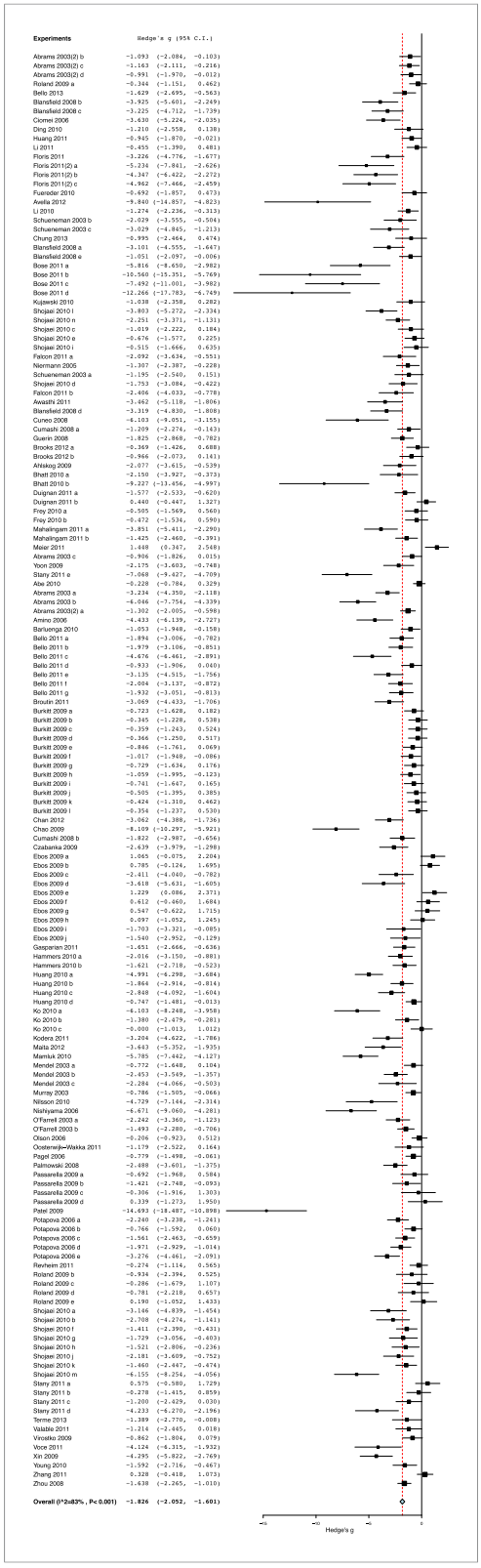

Figure 2—figure supplement 1. Effect sizes for all included experiments (n = 158).
DOI: 10.7554/eLife.08351.010

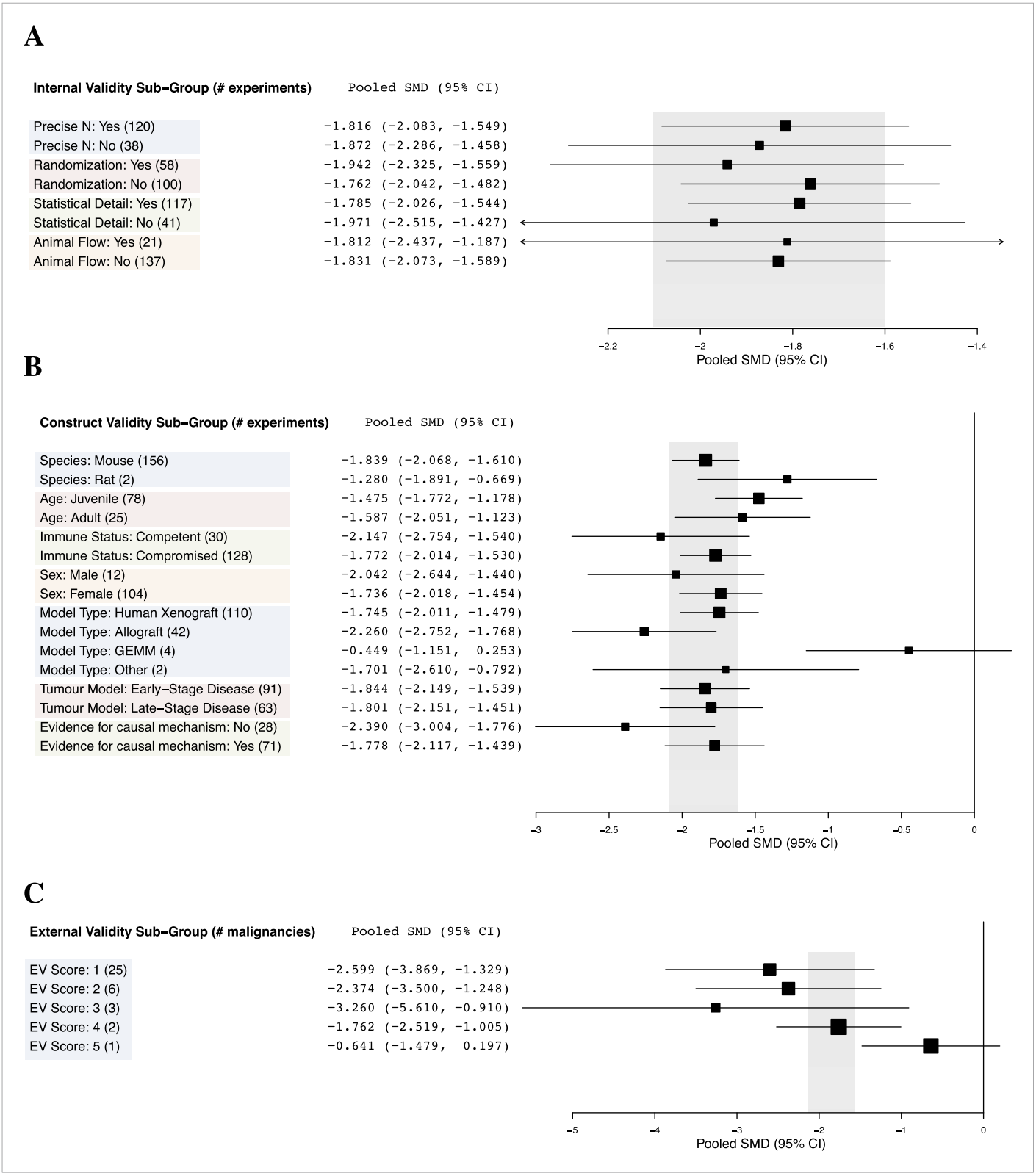

Figure 3. Relationship between study design elements and effect sizes. The shaded region denotes the pooled SMD and 95% CI (−1.8 [−2.1, −1.6]) for all experiments combined at the LCT.

DOI: 10.7554/eLife.08351.011

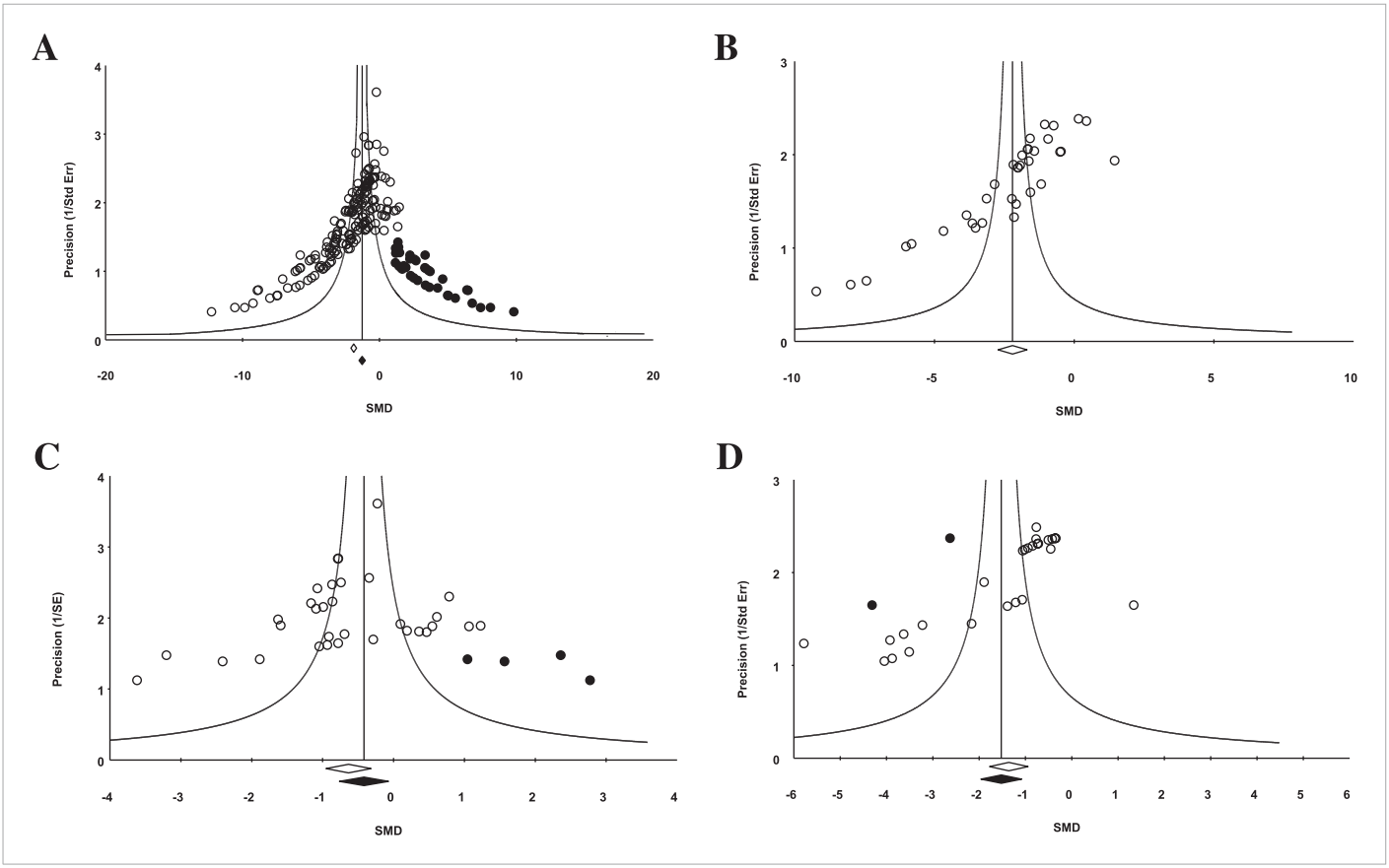

Figure 4. Funnel plot to detect publication bias. Trim and fill analysis was performed on pooled malignancies, as well as the three malignancies with the greatest study volume. (**A**) All experiments for all malignancies (n = 182), (**B**) all experiments within renal cell carcinoma (RCC) (n = 35), (**C**) breast cancer (n = 32), and (**D**) colorectal cancer (n = 29). Time point was the LCT. Open circles denote original data points whereas black circles denote 'filled' experiments. Trim and fill did not produce an estimate in RCC; therefore, no overestimation of effect size could be found.

DOI: 10.7554/eLife.08351.012

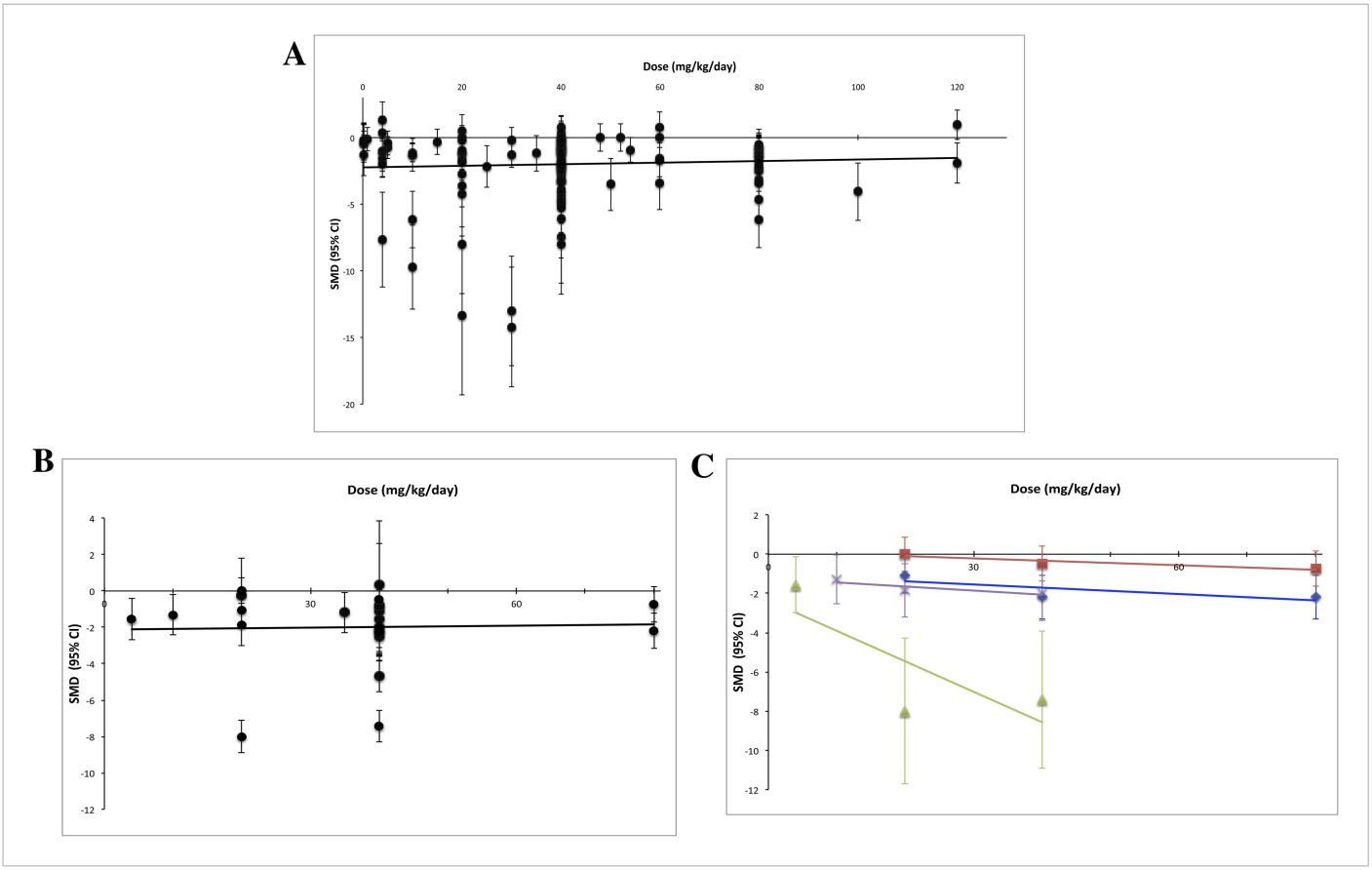

Figure 5. Dose–response curves for sunitinib preclinical studies. Only experiments with a once daily (no breaks) administration schedule were included in both graphs. Effect size data were taken from a standardized time point (14 days after first sunitinib administration). (**A**) Experiments (n = 158) from all malignancies tested failed to show a dose–response relationship. (**B**) A dose–response relationship was not detected for RCC (n = 24). (**C**) Dose–response curves reported in individual studies within the RCC subset showed dose–response patterns (blue diamond = Huang 2010a [n = 3], red square = Huang 2010d [n = 3], green triangle = Ko 2010a [n = 3], purple X = Xin 2009 [n = 3]).

DOI: 10.7554/eLife.08351.013