
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

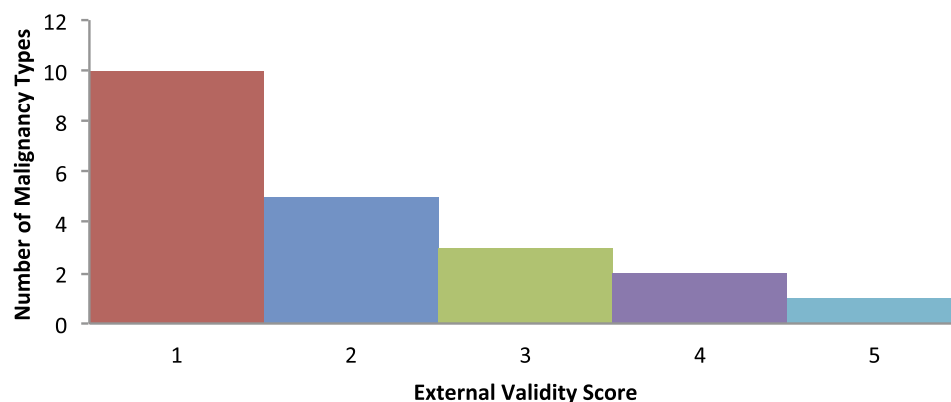


Figure 1. Descriptive analysis of (A) internal, construct, and (B) external validity design elements. External validity scores were calculated for each malignancy type tested, according to the formula: number species used + number of models used; an extra point was assigned if a malignancy type tested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

DOI: [10.7554/eLife.08351.003](https://doi.org/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%)
Addressed 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%)

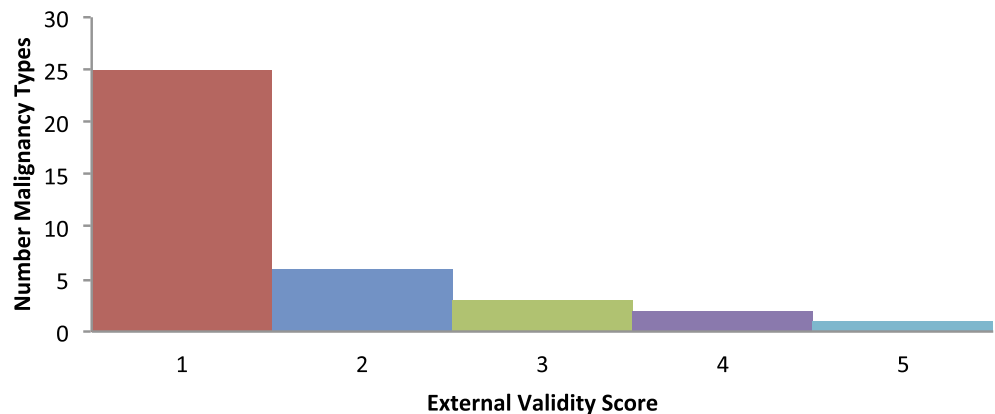
B

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](https://doi.org/10.7554/eLife.08351.005)

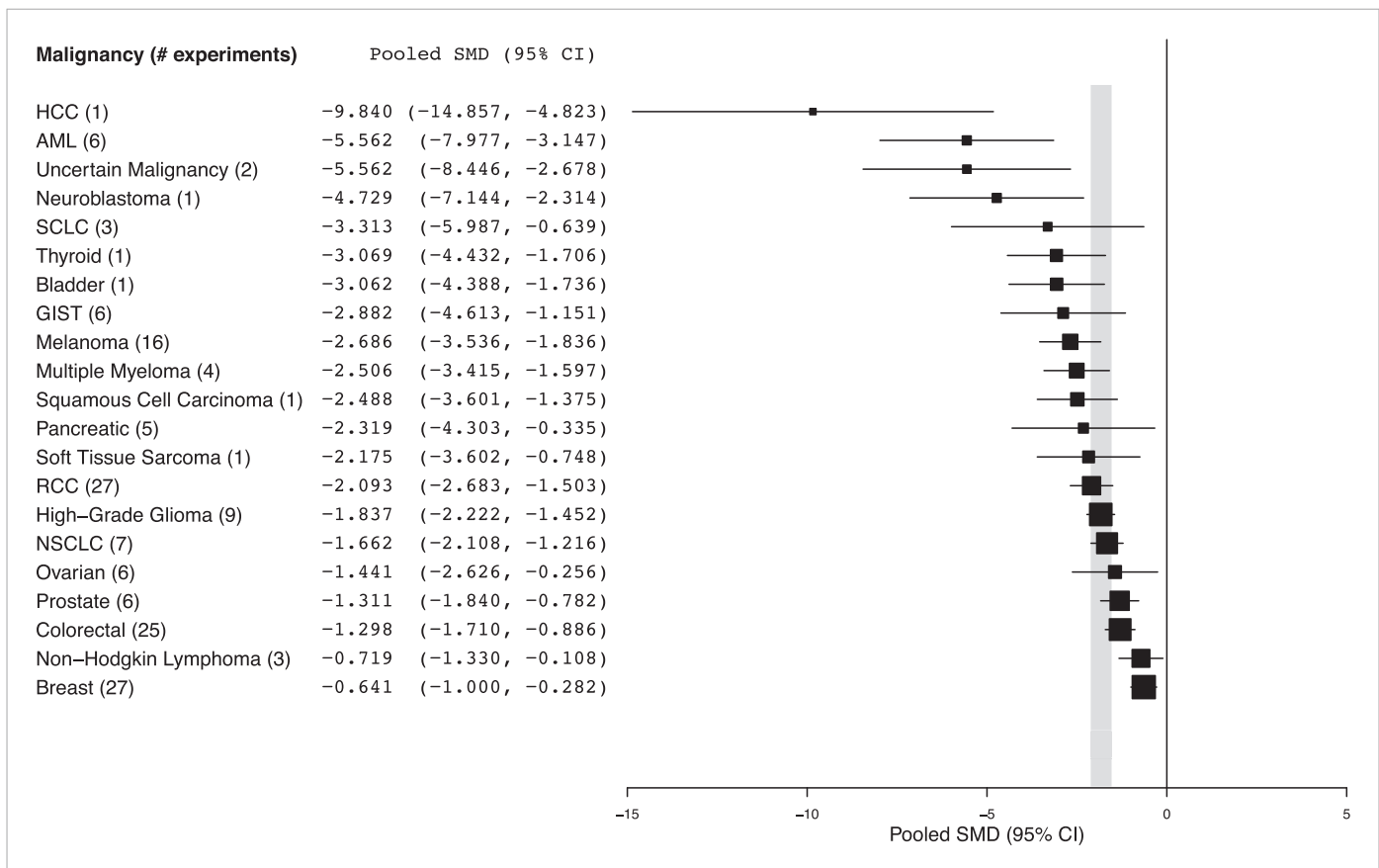


Figure 2. Summary of pooled SMDs for each malignancy type. Shaded region denotes the pooled standardized mean difference (SMD) and 95% confidence interval (CI) (-1.8 [-2.1, -1.6]) for all experiments combined at the last common time point (LCT).

DOI: [10.7554/eLife.08351.008](https://doi.org/10.7554/eLife.08351.008)

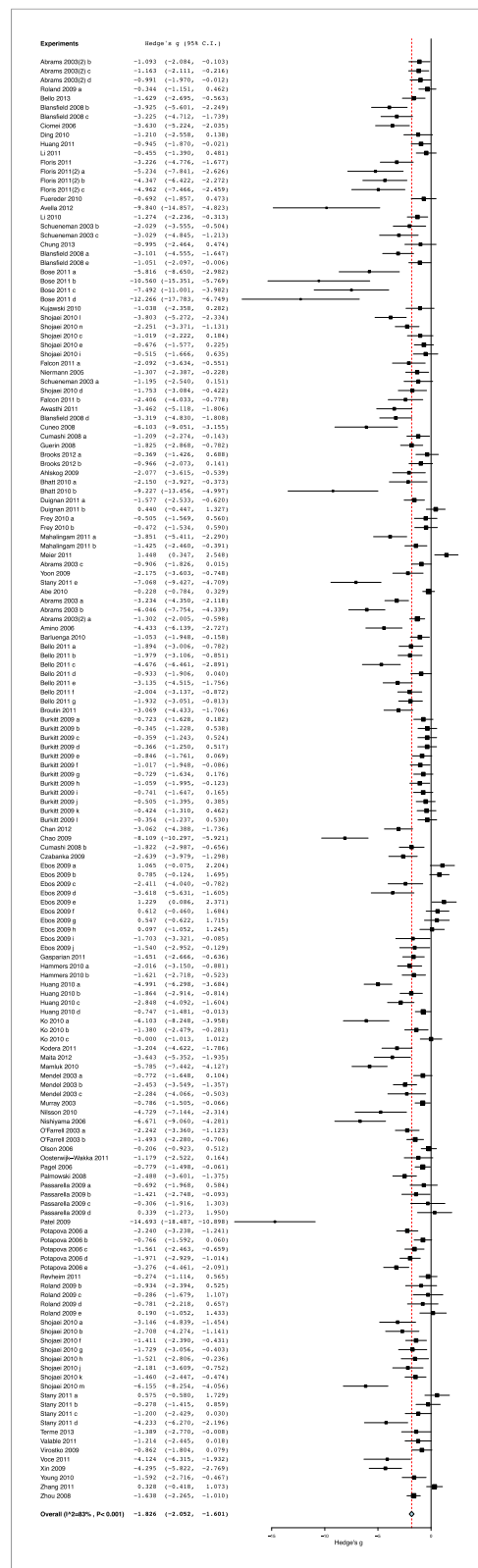
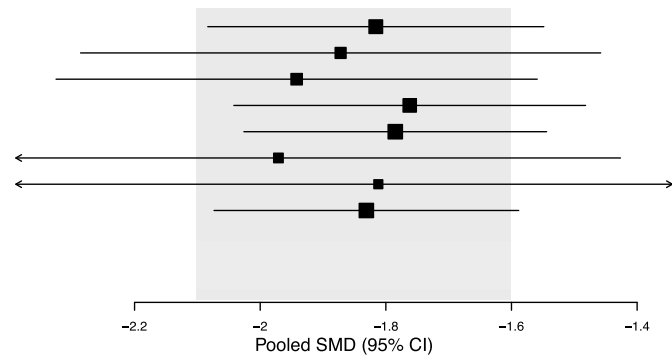


Figure 2—figure supplement 1. Effect sizes for all included experiments (n = 158).

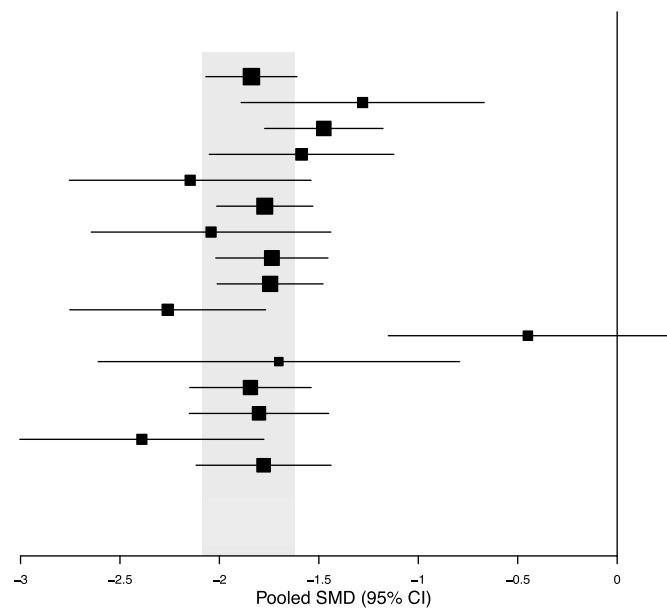
DOI: [10.7554/eLife.08351.010](https://doi.org/10.7554/eLife.08351.010)

A**Internal Validity Sub-Group (# experiments)** Pooled SMD (95% CI)

Precise N: Yes (120)	-1.816 (-2.083, -1.549)
Precise N: No (38)	-1.872 (-2.286, -1.458)
Randomization: Yes (58)	-1.942 (-2.325, -1.559)
Randomization: No (100)	-1.762 (-2.042, -1.482)
Statistical Detail: Yes (117)	-1.785 (-2.026, -1.544)
Statistical Detail: No (41)	-1.971 (-2.515, -1.427)
Animal Flow: Yes (21)	-1.812 (-2.437, -1.187)
Animal Flow: No (137)	-1.831 (-2.073, -1.589)

**B****Construct Validity Sub-Group (# experiments)** Pooled SMD (95% CI)

Species: Mouse (156)	-1.839 (-2.068, -1.610)
Species: Rat (2)	-1.280 (-1.891, -0.669)
Age: Juvenile (78)	-1.475 (-1.772, -1.178)
Age: Adult (25)	-1.587 (-2.051, -1.123)
Immune Status: Competent (30)	-2.147 (-2.754, -1.540)
Immune Status: Compromised (128)	-1.772 (-2.014, -1.530)
Sex: Male (12)	-2.042 (-2.644, -1.440)
Sex: Female (104)	-1.736 (-2.018, -1.454)
Model Type: Human Xenograft (110)	-1.745 (-2.011, -1.479)
Model Type: Allograft (42)	-2.260 (-2.752, -1.768)
Model Type: GEMM (4)	-0.449 (-1.151, 0.253)
Model Type: Other (2)	-1.701 (-2.610, -0.792)
Tumour Model: Early-Stage Disease (91)	-1.844 (-2.149, -1.539)
Tumour Model: Late-Stage Disease (63)	-1.801 (-2.151, -1.451)
Evidence for causal mechanism: No (28)	-2.390 (-3.004, -1.776)
Evidence for causal mechanism: Yes (71)	-1.778 (-2.117, -1.439)

**C****External Validity Sub-Group (# malignancies)** Pooled SMD (95% CI)

EV Score: 1 (25)	-2.599 (-3.869, -1.329)
EV Score: 2 (6)	-2.374 (-3.500, -1.248)
EV Score: 3 (3)	-3.260 (-5.610, -0.910)
EV Score: 4 (2)	-1.762 (-2.519, -1.005)
EV Score: 5 (1)	-0.641 (-1.479, 0.197)

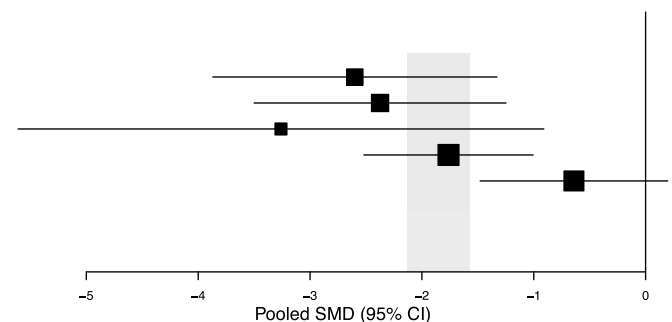


Figure 3. Relationship between study design elements and effect sizes. The shaded region denotes the pooled SMD and 95% CI (-1.8 [-2.1, -1.6]) for all experiments combined at the LCT.

DOI: [10.7554/eLife.08351.011](https://doi.org/10.7554/eLife.08351.011)

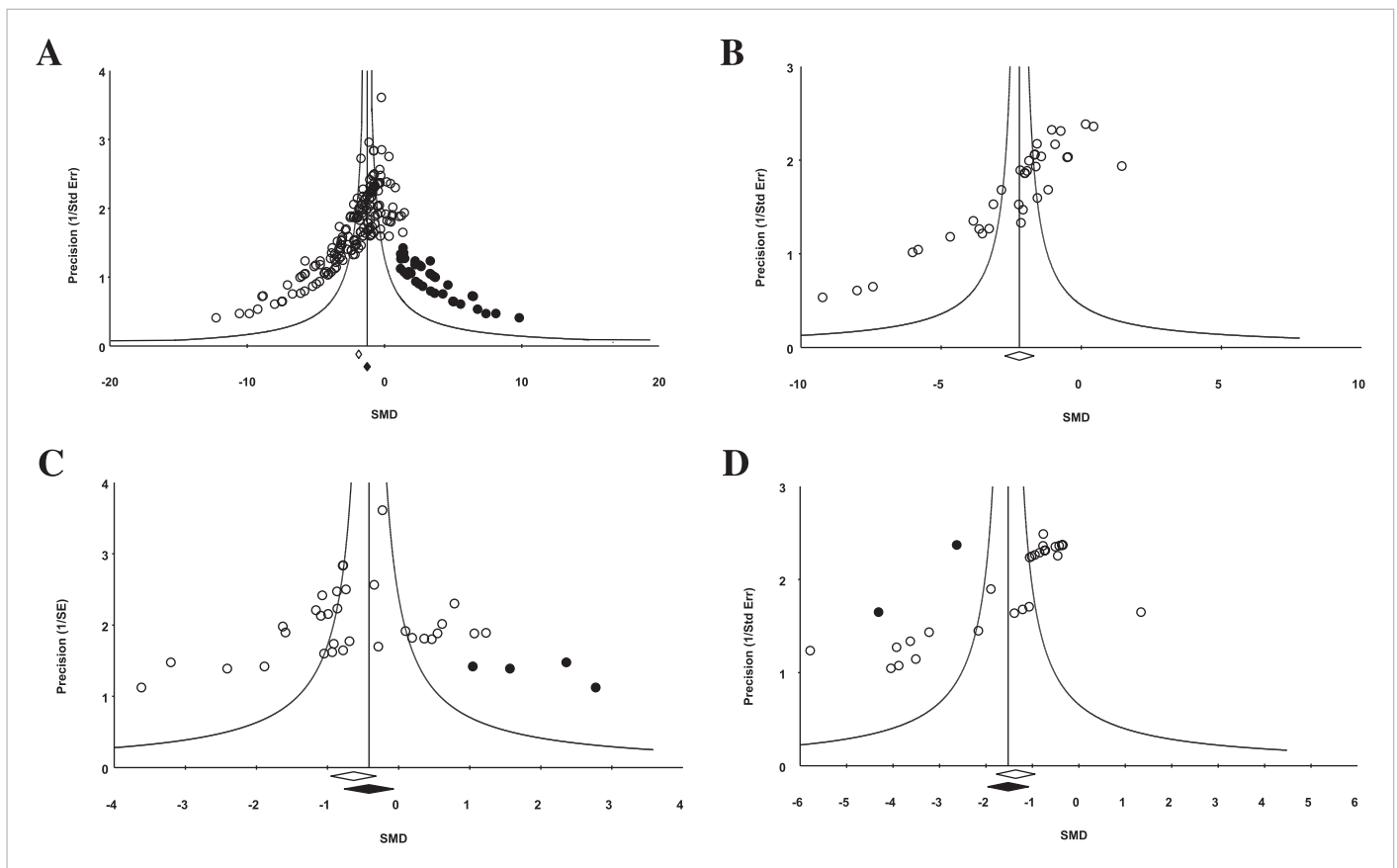


Figure 4. Funnel plot to detect publication bias. Trim and fill analysis was performed on pooled malignancies, as well as the three malignancies with the greatest study volume. **(A)** All experiments for all malignancies ($n = 182$), **(B)** all experiments within renal cell carcinoma (RCC) ($n = 35$), **(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](https://doi.org/10.7554/eLife.08351.012)

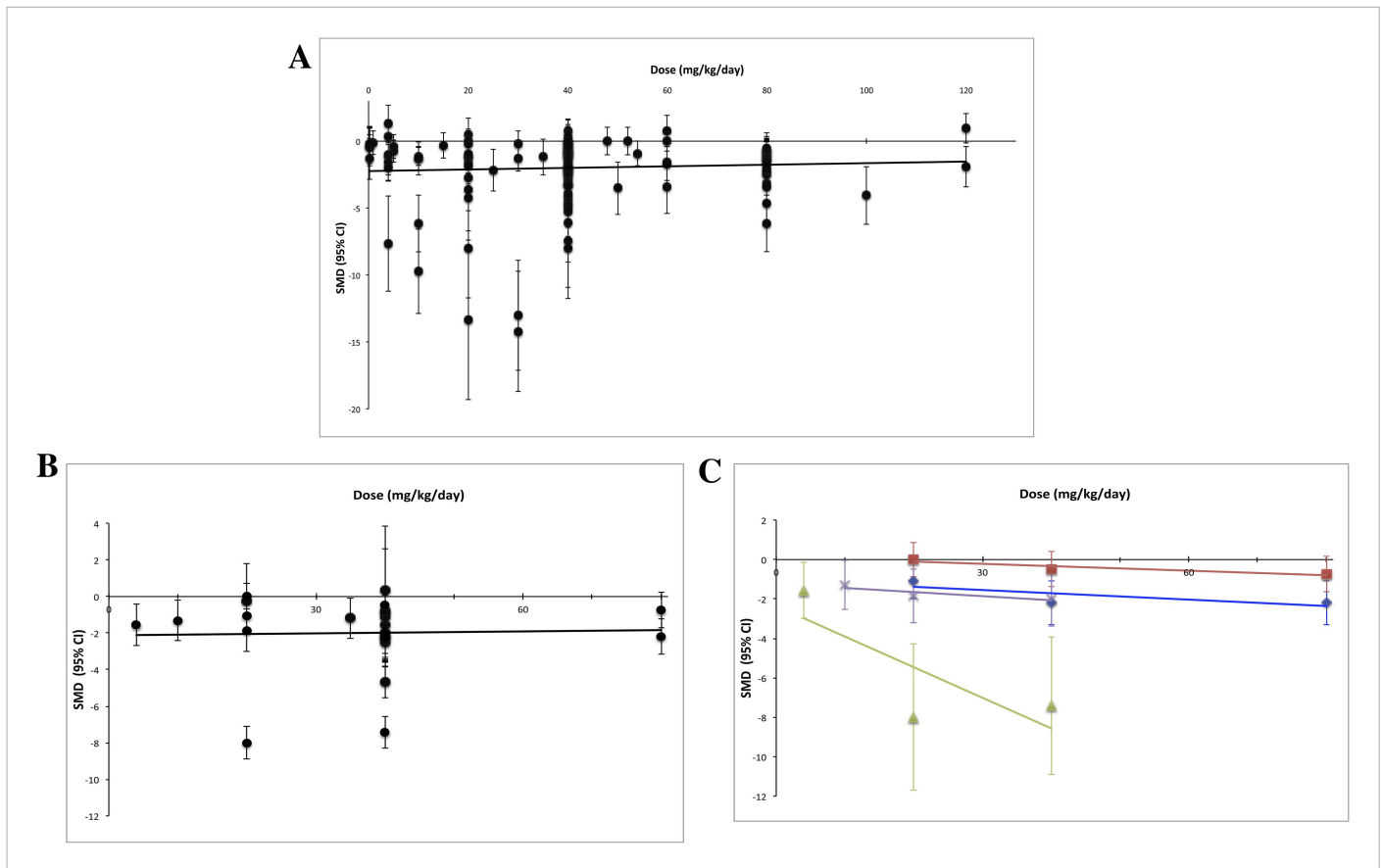


Figure 5. Dose–response curves for sunitinib preclinical studies. Only experiments with a once daily (no breaks) administration schedule were included in both graphs. Effect size data were taken from a standardized time point (14 days after first sunitinib administration). **(A)** Experiments ($n = 158$) 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](https://doi.org/10.7554/eLife.08351.013)