Supplementary Material

Logistic Regression

Methods

To identify variables with a strong effect on the probability of publication in an exploratory analysis, we used a logistic regression model to test several explanatory variables. The binary outcome of the logistic regression was a timely publication (defined as posting results on ClinicalTrials.gov within **1 year** or/and publishing journal article within **2 years**). We tested the following variables: industry (yes/no), medication (yes/no), completion years (2009-2013), phase (I-IV), number of participants, number of centers (mono-/multicenter) and population (adult/pediatric/mixed/unclear).

To identify the relevant explanatory variables, the logistic regression model was built stepwise, starting with univariate models and adding variables in each step. A log-likelihood ratio was used in each step to compare the more complex model (with an additional variable) to the less complex. In each step the variable that led to the largest log-likelihood ratio was included in the subsequent model until adding subsequent variables did not improve the model further. As this was an exploratory analysis, we had not defined strict rules for variable inclusion a priori, not did we consider interaction terms.

Results

Table S1: Results for the univariate logistic regression models with each of the possible explanatory variables added to the model.

Variable	Odds ratio (95% CI)	Log-likelihood score
Constant	-	-209.2
Industry sponsor	0.62 (0.37 – 1.03)	-207.5
Medication	0.67 (0.41 – 1.09)	-207.8
Completion Year (Odds ratio w.r.t. 2009)		-207.2
2009	1 (ref)	
2010	1.86 (0.84 – 4.12)	
2011	1.43 (0.66 – 3.11)	
2012	1.96 (0.93 – 4.13)	
2013	1.54 (0.76 – 3.14)	
Phase (Odds ratio w.r.t. Phase I)	1 (ref)	-193.5
I-II	0.00 (0.00 – Inf.)	
II	0.32 (0.06 – 1.70)	
II-III	2.25 (0.31 – 16.41)	
III	1.22 (0.24 – 6.28)	
IV	1.62 (0.29 – 9.05)	
No phase (Non-drug trials)	0.73 (0.13 – 4.08)	
Number of participants	1.00075 (1.00016 – 1.0013) (Increase in odds per participant, increase by 500 participants increases odds by 1.45)	-204.3
Multicenter study	0.64 (0.36 – 1.14)	-208.0
Population (Odds ratio w.r.t. adult)		-208.3

Adult	1 (ref)	
Pediatric	0.96 (0.53 – 1.74)	
Mixed	2.32 (0.21 – 25.88)	
Unclear	0.00 (0.00 – Inf.)	

As presented in table S1, for the univariate model the variable 'Phase' clearly led to the highest increase in the log-likelihood score. The second-highest log-likelihood was reached through the 'number of participants' variable.

Using the stepwise procedure, the following variables were added to the model: 'Phase' (resulting in log-lik. -193.5, compared to -209.2 for the constant model) and then 'Enrollment' (log-lik. -190.7). While the strongest decrease in likelihood appeared in the first step, the next selection step yielded only a smaller decrease. Adding the next variable (i.e. the 'Completion year' variable) did not relevantly improve the log likelihood score (log-lik. -188.9). The area under the curve, which is a measure for the predictiveness of the model, yielded a value of AUC = 0.686 (95% CI 0.627 – 0.745) for the final model, which is a rough indication that the selected variables had some predictive power.