Box S1 | Data collection and analysis

All data used in this analysis was collected directly from 7 major sponsor companies, all of which are in the top 20 biopharma companies ranked by revenue in 2016. Standardized definitions and methodology were utilized to determine the costs for each trial. All data went through numerous quality checks to ensure that companies were properly reporting the figures and adhering to the strict criteria and definitions provided.

The analysis is based on 726 interventional patient studies with costs allocated over the 2010–2015 timeframe.

Trials by phase	
Phase	Trials (*N*)
Phase I (patients)	189
Phase II	264
Phase III	273
Total	726

The dataset encompassed two types of data: Trial Details Data and Clinical Financial Data. The Trials Detail Data included information about each trial active in the dataset: for example, disease, key milestones, phase, as well as detailed information on each site initiated for the trial. This enabled KMR Group to analyze certain characteristics of the trials and assess cost drivers such as number of sites, subjects, duration, geography and outsourcing use.

The Clinical Financial Dataset was pulled together with the help of finance groups from each of the participants. Since companies use different mechanisms and methods to track trial cost, a robust top-down approach was agreed upon to ensure comparability across companies.

Clinical development spending was collected on an annualized basis over 6 years by cost type according to specific inclusion criteria. This method ensured equivalence across companies as to the cost basis including types of costs included in the assessment and eliminated the bias of some companies including fewer (for example, only cost centres directly linked to trials) or differing types of cost centres (for example, drug supply). It also enabled KMR Group to validate the cost centres. This sum was the basis for assigning costs to trials.

The ability to allocate consolidated annual spending to individual active trials was critical to the success of this study.

Direct costs were assigned automatically to trials, since these are systematically tracked on a trial basis in most biopharma company financial systems. Examples of direct costs include investigator grants and contract research organization (CRO) costs. Personnel costs were assigned to trials using staff hours tracked to trials. Hours were gathered from detailed time reporting systems, which were then used to pro-rate personnel spending to individual trials.

The analysis relating to completed trials was limited to those with a final clinical trial report (CTR) date in 2012–2015 and that began 2010 or later. This window met the limits of financial and time reporting systems and still captured full costs of the trial. The full dataset, including active trials that were not closed, was used for other analyses (for example, burn rates), which was made available in a full report and online in the Trial Cost Metrix application.

This top-down approach and cost assignment method ensured that all costs relating to a trial were covered, including costs that were incurred prior to protocol approval (for example, planning-related expenses) as well as after CTR.

Costs for all years were adjusted for inflation to 2015.

Statistical analysis

Given the depth and detail of the data collected, numerous statistical approaches were used to provide insights to participants.

Single variable analysis, through regressions and chi-squared tests, were applied to test how certain trial characteristics relate to cost (for example, number of subjects, work in emerging markets, molecule size), for both statistical significance and magnitude of effect. This was used to determine how certain factors contribute to overall trial costs.

The following table (presented as Table 1 in the main article) summarizes the statistical significance of several factors on trial cost:

Factor	*p* value
Sites	<0.0001
Subjects	<0.0001
Visits	<0.0001
Duration	<0.0001
Molecule size	0.29237
Rare disease	0.16636
Adaptive design	0.24437
Emerging market activity	<0.0001
Emerging market subjects	0.00013
Regions	<0.0001
Countries	<0.0001

Once these contributing factors were determined, we further investigated the interactions between variables to understand what matters most for trial cost. Correlations, multivariate modelling, and analysis of data subsets were used to assess the presence of confounding variables in these results. This analysis formed the basis for the construction of a multivariate model.

To rank operational performance for companies, a multivariate model was constructed that normalized for trial design parameters. Using a multivariate regression with volume of sites, therapy area, and treatment duration as the independent variables, all data was put through the model equation and the residual (that is, actual spending minus model-derived spending) was calculated for each trial. Each company’s performance was then ranked on the magnitude of this residual: those spending less than the model-derived value were high performers while those that spent more were low performers. After identifying which companies were high and low performers, we were able to run correlations across companies, investigating certain relationships such as: do companies that expend a higher proportion of clinical development on outsourcing tend to reap operational benefits in their performance?