Unsuccessful Trial Accrual and Human Subjects Protections: An Empirical Analysis of Recently Closed Trials

Benjamin Carlisle, MA[1], **Jonathan Kimmelman, PhD**[1], **Tim Ramsay, PhD**[2], and **Nathalie MacKinnon, BSc**[1]

[1]Studies of Translation, Ethics and Medicine (STREAM), Biomedical Ethics Unit, McGill University, 3647 Peel Street, Montréal QC H3A 1X1

[2]Ottawa Hospital Research Institute, Ottawa, Canada

Abstract

BACKGROUND—Ethical evaluation of risk/benefit in clinical trials is premised on the achievability of resolving research questions motivating an investigation.

OBJECTIVE—To determine the fraction and number of patients enrolled in trials that were at risk of not meaningfully addressing their primary research objective due to unsuccessful patient accrual.

METHODS—We used the National Library of Medicine clinical trial registry to capture all initiated phase 2 and 3 intervention clinical trials that were registered as closed in 2011. We then determined the number that had been terminated due to unsuccessful accrual and the number that had closed after less than 85% of the target number of human subjects had been enrolled. Five factors were tested for association with unsuccessful accrual.

RESULTS—Of 2579 eligible trials, 481 (19%) either terminated for failed accrual or completed with less than 85% expected enrolment, seriously compromising their statistical power. Factors associated with unsuccessful accrual included greater number of eligibility criteria (p=0.013), non-industry funding (25% vs. 16%, p <0.0001), earlier trial phase (23% vs. 16%, p <0.0001), fewer number of research sites at trial completion (p <0.0001) and at registration (p<0.0001), and an active (non-placebo) comparator (23% vs. 16%, p <0.001).

CONCLUSION—48,027 patients had enrolled in trials closed in 2011 that were unable to answer the primary research question meaningfully. Ethics bodies, investigators, and data monitoring committees should carefully scrutinize trial design, recruitment plans, and feasibility of achieving accrual targets when designing and reviewing trials, monitor accrual once initiated, and take corrective action when accrual is lagging.

Correspondence to: Jonathan Kimmelman.

Introduction

All major policies of research ethics require a favourable balance of risks against benefits to human subjects, if any, and to society. Investigators and local ethics committees make this determination at the outset of clinical investigations, and data and safety monitoring committees are instructed to ensure that an acceptable risk-benefit balance is maintained over the course of a study.[3]

Risk-benefit assessments typically focus on specific design elements in a protocol, including study drugs or other interventions, research procedures, choice of comparator, and inclusion criteria. In all cases, however, a trial's risk-benefit balance is premised on a study being implemented as described in a protocol, and trials that encounter insurmountable barriers to execution expose human subjects to research burdens without compensatory gains in medical knowledge. One particularly common impediment to successful trial execution is poor recruitment. An Institute of Medicine (IOM) report cited 71% of phase 3 trials approved by the National Cancer Institute's (NCI's) Cancer Therapy Evaluation Program (CTEP) closed without meeting 100% of their accrual goals.[4] A subsequent IOM report indicated that 40% or more NCI-sponsored phase 3 trials failed to meet accrual goals.[5] An NCI study of 149 trials estimated that 28.3% would fail to achieve 90% of their accrual goals.[6] Louis Lasagna famously observed that when trial recruitment starts, "the supply of suitable patients becomes a fraction of what it was assumed to be before the trial began."[7]

Failure to enrol and retain a target sample reduces a study's statistical power, compromising its prospect of delivering a statistically informative answer to the primary question grounding its design and review. Abortive studies also poorly utilize scarce human and material resources, and deplete the supply of eligible candidates for other investigations.[8–10]

Accrual effectiveness once a trial is initiated thus has important implications for safeguarding the welfare of human subjects and the broader research enterprise. We used the clinical trial register clinicaltrials.gov to determine the volume of initiated trials that are unable to attain a meaningful sample, to examine the relationship between unsuccessful accrual and trial closure, and to identify factors that may confer risk of unsuccessful accrual.

Methods

Our primary goal was to determine the volume of trials that were unable meaningfully to address the primary research question due to inadequate patient accrual. Secondarily, we evaluated several factors for association with accrual failure and examined the relationship between trial termination and completion with unsuccessful accrual.

We began by downloading all records of trials in which more than one human subject had enrolled and that ended in year 2011 from the National Library of Medicine (NLM) trial register, clinicaltrials.gov. We used two different queries. First, we searched for all trials marked as "completed" and included all trials with an end date in year 2011. NLM defines "completed" as "stud[ies that] ha[ve] ended normally, and participants are no longer being examined or treated." Second, we searched for all trials listed as "terminated" with an end

date in the year 2011. NLM defines "terminated" as studies that "stopped recruiting or enrolling participants early and will not start again."

For trials captured in the first query, we defined trials with "unsuccessful accrual" as those in which more than one human subject had enrolled but enrolment at completion was less than 85% of expected enrolment based on the initial trial entry. This threshold, which was established *a priori*, was chosen as a reasonable figure at which the statistical power for the primary endpoint becomes seriously compromised; it is consistent with cut-points used in prior studies of recruitment.

For the second query, we defined trials with "unsuccessful accrual" as those in which more than one human subject had enrolled but for which accrual problems appeared among the reasons for termination. Because reasons for termination are reported in an open text field, they were classified by two independent coders. Agreement between the coders using Cohen's kappa was 0.96.

For both queries, we applied the following inclusion criteria: 1) intervention studies, and 2) at least one patient must have enrolled. We also excluded trial types where power calculations are often not critical for design and interpretation, namely: 1) phase 1 and 4, 2) "hybrid" phase (e.g. "phase 2 / 3"), 3) open label extension and 4) observational. Results of both query methods were combined to create our sample of trials that were unsuccessful due to inadequate accrual. To establish the face validity of our sample, we randomly inspected 50 trial entries from the first query. Of these, only 4% of the entries reflected study designs where sample sizes would be impossible or inappropriate to define *a priori* (e.g. adaptive designs).

We extracted the following data elements from all trial entries: a) lead sponsor's agency class (industry or non-industry), b) the number of research sites at first registration and closure, c) number of eligibility criteria (operationalized as lines in the eligibility criteria), d) whether the study involved placebo comparators, e) phase, f) number of patients enrolled and the anticipated enrolment from the first registration, g) whether the study included paediatric human subjects (operationalized as human subjects under 18).

Trials for which the ratio of actual enrolment at completion to expected enrolment at first registration was less than 0.85 were added to the set of trials terminated for poor accrual; these were compared with the remaining trials completed in 2011 with at least 85% of expected enrolment.

At the outset of our study, we identified five factors that we believed may confer risk of unsuccessful accrual: earlier study phase, non-industry lead sponsor, smaller number of research sites,[12] more eligibility criteria, and use of placebo comparators. Our hypothesis was that trials in earlier stages of development and trials pursued without the resources of an industry lead sponsor would have a harder time recruiting patients. Upon collecting the data, we discovered that the number of sites at registration was often different than the number of sites at study termination; we elected to look at these two factors separately. We also noticed that 419 trials listed having 0 sites at registration and 126 still listed having 0 sites at completion; we excluded these entries from our analysis.

Statistical methods

To test the univariate relationship between each factor individually and successful accrual, we conducted a simple chi-square test on the full dataset. We log-transformed the number of sites, as these variables were very skewed. To test the relationship between these factors and successful accrual, we first tested each factor separately and then used backwards selection to construct a multivariable logistic regression model.

For the multivariable model, we hoped that we could derive a model that would prove useful for predicting which trials were likely to be unsuccessful. Given that automatic model building algorithms are known to overfit, we randomly split the data into training and testing sets of equal size. We used backwards selection to arrive at our final model, and then tested that modns in the independent testing set. We summarized the associations by numerically listing the percent successful accrual in each of the categories defined by the model, as well as with the model C-statistics, which in models with continuous predictors, is equivalent to the area under the ROC curve. As a post hoc test, we used our entire sample to test whether patient enrolment targets showed any relationship with accrual failure. Before performing the test, we defined statistical significance as $p<0.05$.

Results

Volume of Accrual Failures

The flow of trials captured in our search is shown in Figure 1. We identified a total of 2577 eligible phase 2 or 3 trials in which human subjects had enrolled and either were completed or terminated in 2011. Of these, 326 (13%) were classified as completed but failed accrual by our definition.

An additional 363 trials were terminated early for various reasons (Table 1), 72 (20%) of which were "informative terminations, i.e. they were stopped for futility, efficacy, or safety. A total of 15,380 human subjects had enrolled in studies that were terminated prematurely for uninformative reasons; poor accrual accounted for 156 (43%) of all terminated trials.

In total, investigators of 481 (19%) trials, in which 48,027 patients enrolled, were at risk of being unable meaningfully to address their primary research question due to inadequate sample size at study completion or termination resulting from recruitment failure. These trials accounted for 7% of all human subjects in trials that closed in 2011. Of trials with poor accrual, 62 (13%, with a combined enrolment of 8,727 human subjects) included children.

Table 2 presents the number of trials (and participants enrolled) that achieved other fractions of projected enrolment. For example, had we chosen to define unsuccessful accrual as a failure to achieve 100% of expected enrolment (the same cut-off used by Cheng et al[4]) rather than our actual 85% cut-off, 676 trials, representing 26% of our identified eligible trials and 146,157 patients would fall below this threshold.

To examine the timeliness of mechanisms for terminating trials at risk of accrual failure, we plotted the number of trials terminated due to accrual and trials completed against the fraction of expected enrolment (Figure 2). As indicated, the number of trials registered as

completed but that had achieved only 30–40% of their target enrolment exceeded by more than twofold the number of trials registered as terminated after having achieved that same enrolment fraction. One possible explanation for this pattern is that investigators who encountered recruitment problems discontinued recruiting new patients, but kept the trial registered as "open" because they were gathering data on outcomes. To explore this possibility, we examined a random sample of 100 registration records (50 for each group) for the proportion of time between initiation and closure for which trials were registered as open for recruitment. On average, trials that attained <85% their target sample were open for recruitment for a greater proportion of the total time open than trials that attained >85% their target sample (78% vs. 67%). In addition, trials that had attained <85% showed a longer mean total period from initiation and closure (156 vs. 135 weeks).

Risk Factors Associated with Accrual Failure

Each of the five different potential risk factors for incomplete enrolment was significantly associated with the incomplete enrolment, but only 4 as predicted: 23% of Phase 2 trials failed to accrue adequately, versus 16% for Phase 3 ($p<0.0001$); a greater number of eligibility criteria was associated with inadequate accrual ($p=0.0125$); 16% of placebo-controlled trials failed to accrue adequately compared to 23% for active control ($p<0.0001$); 16% of industry-funded trials failed to adequately accrue adequately compared to 25% for publicly-funded trials ($p<0.0001$); a larger number of trial sites was associated with better accrual, regardless of whether the number analyzed was recorded at trial registration or completion ($p<0.0001$).

To develop a set of factors that could be used to flag trials unlikely to accrue successfully, we used a backwards selection strategy to derive a multivariable logistic regression model with incomplete enrolment as the dependent variable. The resulting model included only trial phase and source of funding. The area under the ROC curve for this model was only 0.59 in the validation set, a number that reflects weak associative power as an area under the curve of 0.5 indicates that the model performed no better than a guess. Table 3 presents the percentage of trials completed in 2011 that had incomplete enrolment in each of the four categories defined by the final model. In the highest risk category, publicly-funded phase 2 trials, 78% of trials achieved adequate enrolment. Post hoc, we found that larger expected enrolment correlated with recruitment success ($p<0.00001$).

Discussion

We report that 19% of trials registered as newly closed in 2011 either terminated due to failed accrual or completed with less than 85% of their expected enrolment, thus likely decreasing their statistical power below that planned at trial initiation. To our knowledge, this investigation represents the most comprehensive study of accrual failure.[4–6] Our findings, though troubling, paint a less grim picture than if we had chosen a cut-off for unsuccessful accrual of 100%; in that case, the fraction of trials failing to achieve enrolment targets would have risen to 29%, much lower than the fraction reported by the NCI in 2010.[13]

We further explored factors associated with failure to achieve sufficient accrual to address primary research questions. As expected, number of eligibility criteria, non-industry sponsorship, earlier trial phase, and fewer study centres all were associated with failed accrual. We were surprised to find that trials with active controls more often failed accrual than placebo-controlled trials after considering all factors evaluated. Since smaller trials are at greater risk of accrual failure, proportionally fewer human subjects overall (7%) are enrolled in trials that are unable to accrue successfully. However, none of the variables we examined, either alone or in combination with others, was associated strongly with accrual failure.

Unsuccessful accrual typically is viewed as a practical problem for trials and medical centres. Here, we examine some moral dimensions of the issue. Risk-benefit favourability in trials generally is premised on burdens endured by volunteers being redeemed by addressing the principal questions driving a study. A favourable risk-benefit balance at outset is diminished by factors that impede trial execution. Human subjects who enrol early on in trials that later stall in recruitment are exposed to a risk-benefit ratio that is eroded from that described in the protocol. Obviously, even studies that fail to approach their recruitment target still return some information: the moral problem of under-accrual may be somewhat mitigated by synthesis of findings in subsequent meta-analyses,[15] and secondary endpoints and safety observations recorded are still informative. Just the fact that a study has less statistical power than planned does not necessarily imply that the risk-benefit balance has fallen below the threshold for moral acceptability.

Nevertheless, to maintain a risk-benefit balance established per protocol, trials encountering serious recruitment problems should be considered for termination. The number of trials that are terminated early for failed accrual ideally should exceed the number of trials that are continued to completion but with enrolments that are far below the target sample. Our findings suggest that researchers tend to persevere in recruiting patients even when meaningful accrual is futile. Thus, investigators who encounter recruitment problems terminate recruitment short of their goal, but nevertheless complete follow up of human subjects already enrolled for outcome assessment and possibly ancillary information. However, our finding that low accrual trials, on average, run longer and for a greater proportion of their time are open for recruitment does not support this explanation.

We believe our findings have implications for trial design, planning, review, and monitoring. First and most obviously, researchers should develop better systems for attracting eligible patients to well-designed trials and encouraging them to participate. When designing a trial, investigators and review committees should consider whether aspects of a trial's design would impede adequate enrolment. Exclusion criteria that are too narrow, while seeming to protect the safety of human subjects, can actually diminish a trial's risk-benefit balance, as may excessive demands on patients. Second, our finding that almost a fifth of trials completed in 2011 failed to meet 85% of accrual goals suggests a significant level of inefficiency in the clinical research enterprise. Institutional review boards, ethics committees, and investigators are instructed to assess and evaluate risk-benefit systematically at trial outset. Assessment of operational feasibility should be integral to that evaluation; thus, proposing investigators should be expected to marshal evidence to support

the feasibility of achieving accrual targets. Third, our findings suggest that mechanisms for terminating trials are insufficiently sensitive to recruitment futility. Trial protocols should describe and implement more effective mechanisms for monitoring recruitment futility, re-evaluating risk-benefit, and terminating trials if necessary. Last, our findings provide a starting point for developing indices that alert sponsors and investigators of high risk of recruitment failure. We identified several factors that correlated with recruitment failure. Though we were unable to use our data to develop a useful decision tool for identifying studies at high risk of failed accrual, a prospective study that examined additional data on study characteristics, recruitment practices, or recruitment trends within the first year of a study should permit development of indices that would provide investigators and data monitoring committees an actuarial basis for evaluating risk of failed accrual. We also believe our findings may have implications for informed consent. Specifically, when studies are at risk for recruitment failure, new human subjects as well as human subjects already enrolled should be informed that the study is at risk of being unable to deliver its full promise due to accrual that is slower than originally expected.

Our study has several limitations. First, the NLM clinical trial register contains many erroneous entries.[16] We confirmed that target recruitment numbers at registration are consistent with expected enrolment at first date of enrolment. However, reasons reported for trial termination may be inaccurate, or investigators may have misclassified "terminated" studies as "completed." Second, the factors associated with dropouts that occur during the course of a trial may be different from those associated with poor accrual. Trials indicating that either dropouts or accrual failure were the cause of termination were both included as a part of the set of trials with inadequate accrual, because the ethical considerations with regard to the risk-benefit balance will be the same. Third, our 85% cut-off was an arbitrary choice; not all trials that fail to reach this target are uninformative; the simple fact that a study is underpowered does not make it unethical.[17] For example, trials stopped short of their target number can yield information on safety and can detect large intervention effects. Also, though trials with incomplete enrolment are prone to higher type II error, results can be aggregated with findings from other trials in systematic reviews. Nevertheless, we consider uncontroversial the proposition that trials enrolling fewer than 85% are at high risk for failing to address their primary research objective in a meaningful way.

Finally, our analysis embeds moral premises that require qualification. That a trial's risk-benefit balance has worsened due to inadequate recruitment does not necessarily mean that it has fallen below a threshold of ethical acceptability. A trial with less statistical power than expected may still be ethical, all things considered. Our findings suggest that 7% of patients participate in trials that fail to achieve the risk-benefit balance projected, per protocol, at study outset. We do not wish to imply that this fraction of human subjects was exposed to excess risk because they enrolled in trials with inadequate accrual that thus dropped below a threshold of acceptable risk.

Conclusion

Ineffective accrual practices do not necessarily reflect a moral failing on the part of investigators, sponsors, or oversight committees. Unforeseen circumstances, such as a shift

in standard of care or a natural disaster, can derail careful trial planning. Nevertheless, once trials begin enrolment, investigators and trial overseers should view full accrual as a vehicle for maintaining a favourable risk/benefit balance. Our findings suggest that a sizeable fraction of studies were unable to be completed as planned, resulting in costs for both human subjects and the research enterprise.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. World Medical Association. Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. 2011

2. U.S. Department of Health & Human Services. Code of Federal Regulations: Title 45 - Public Welfare - Part 46 - Protection of Human Subjects, 45 CFR 46. 2009

3. The National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research. Department of Health Education and Welfare. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. 1979.

4. Cheng SK, Dietrich MS, Dilts DM. A sense of urgency: Evaluating the link between clinical trial development time and the accrual performance of cancer therapy evaluation program (NCI-CTEP) sponsored studies. Clinical cancer research: an official journal of the American Association for Cancer Research. 2010; 16:5557–63. [PubMed: 21062929]

5. Scoggins JF, Ramsey SD. A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program. Journal of the National Cancer Institute. 2010; 102:1371.

6. Korn EL, Freidlin B, Mooney M, Abrams JS. Accrual experience of National Cancer Institute Cooperative Group phase III trials activated from 2000 to 2007. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2010; 28:5197–201. [PubMed: 21060029]

7. van der Wouden JC, Blankenstein AH, Huibers MJ, van der Windt DA, Stalman WA, Verhagen AP. Survey among 78 studies showed that Lasagna's law holds in Dutch primary care research. J Clin Epidemiol. 2007; 60:819–24. [PubMed: 17606178]

8. Peters-Lawrence MH, Bell MC, Hsu LL, et al. Clinical trial implementation and recruitment: lessons learned from the early closure of a randomized clinical trial. Contemporary clinical trials. 2012; 33:291–7. [PubMed: 22155024]

9. Torgerson JS, Arlinger K, Kappi M, Sjostrom L. Principles for enhanced recruitment of subjects in a large clinical trial. the XENDOS (XENical in the prevention of Diabetes in Obese Subjects) study experience. Controlled clinical trials. 2001; 22:515–25. [PubMed: 11578785]

10. Malmqvist E, Juth N, Lynoe N, Helgesson G. Early stopping of clinical trials: charting the ethical terrain. Kennedy Institute of Ethics journal. 2011; 21:51–78. [PubMed: 21598846]

11. Schroen AT, Petroni GR, Wang H, et al. Achieving sufficient accrual to address the primary endpoint in phase III clinical trials from U.S. Cooperative Oncology Groups. Clinical cancer research: an official journal of the American Association for Cancer Research. 2012; 18:256–62. [PubMed: 21976533]

12. Viberti G, Slama G, Pozza G, et al. Early closure of European Pimagedine trial. Steering Committee. Safety Committee. Lancet. 1997; 350:214–5. [PubMed: 9250200]

13. Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review. J Clin Epidemiol. 1999; 52:1143–56. [PubMed: 10580777]

14. Lemieux J, Goodwin PJ, Pritchard KI, et al. Identification of cancer care and protocol characteristics associated with recruitment in breast cancer clinical trials. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2008; 26:4458–65. [PubMed: 18802158]

15. Halpern SD, Karlawish JT, Berlin JA. THe continuing unethical conduct of underpowered clinical trials. JAMA. 2002; 288:358–62. [PubMed: 12117401]

16. Prayle AP, Hurley MN, Smyth AR. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. BMJ. 2012; 344:d7373. [PubMed: 22214756]

17. Bacchetti P, McCulloch C, Segal MR. Being 'underpowered' does not make a study unethical. Statistics in Medicine. 2012; 31:4138–9. [PubMed: 22763861]
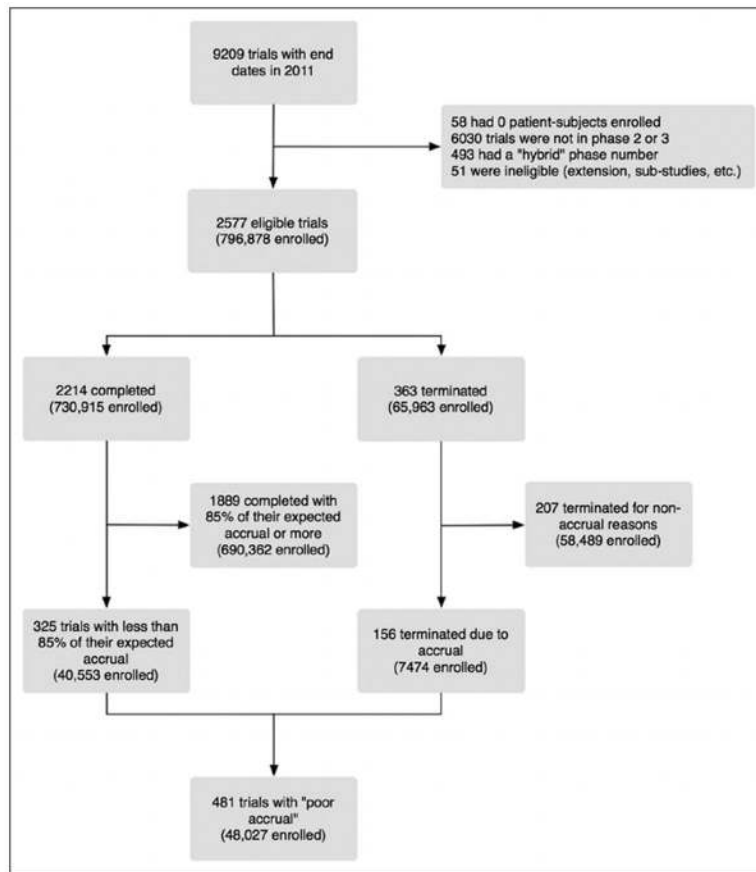
Figure 1.
Trial selection
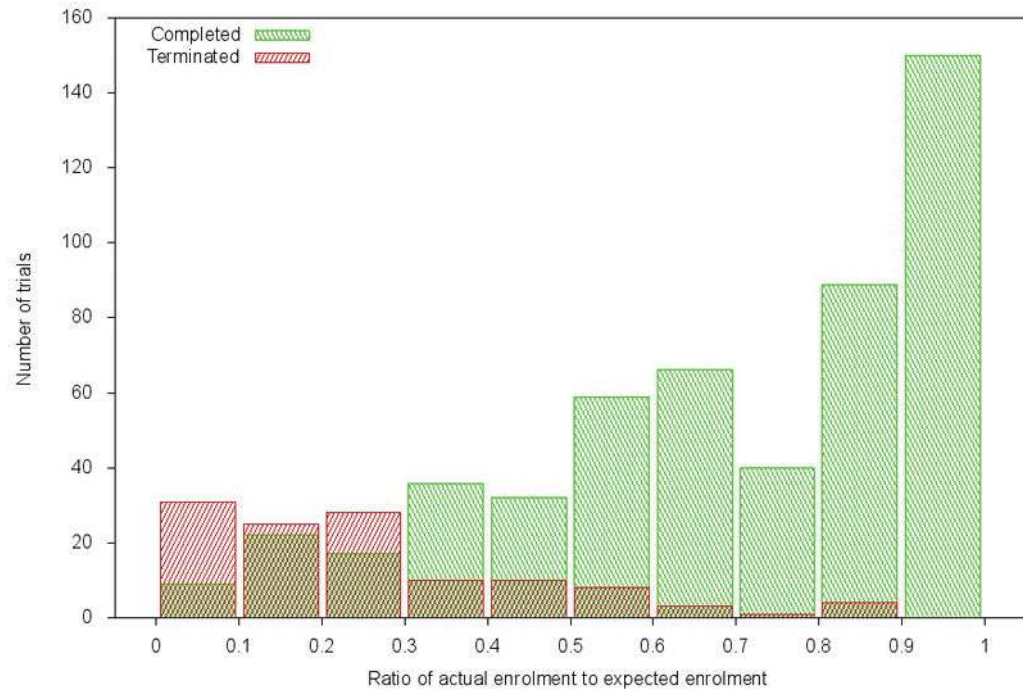
Figure 2.
Ratio of actual enrolment to expected enrolment vs number of trials for trials that completed and trials that terminated due to poor accrual in 2011

Table 1

Cut-off percent of accrual goal vs. proportion of trials that fail to meet it

Cut-off	Number of trials terminated or below cut-off (N=2577)	Number of participants enrolled (N=738,389)
100%	676 (26%)	146,157 (20%)
95%	592 (23%)	78,346 (11%)
90%	526 (20%)	65,943 (9%)
85%	481 (19%)	48,027 (7%)
80%	437 (16%)	38,704 (5%)
75%	415 (16%)	35,773 (5%)
70%	392 (15%)	31,723 (4%)
65%	356 (14%)	27,864 (4%)

Table 2

Reported reasons for the early termination of clinical trials in 2011

Reason for termination	Number of trials	Number of participants enrolled
Poor accrual	156 (43%)	8,504
Informative termination[a]	72 (20%)	30,698
No reason given	65 (18%)	19,885
Funding[a]	31 (9%)	3,598
Administrative[a]	19 (5%)	1,411
"Science moved on"[a]	13 (4%)	633
Unanticipated technical issues[a]	7 (2%)	1,234
All reasons	**363**	**65,963**

[a]See Appendix for reasons included in individual categories.

Table 3

Probability of incomplete enrolment using multivariable logistic regression model.

	Phase 2	Phase 3
Publicly funded	28%	19%
Industry funded	21%	13%