

Chapter 4

Statistical Approaches to Improving Trial Efficiency and Conduct

Janice Pogue, P.J. Devereaux, and Salim Yusuf

Abstract Given the trend towards increasing complexity and cost of clinical trials today, future trials may end up being small, complex, and under-powered to detect clinically meaningful treatment effects. In order to continue to perform important research in the future, we need to make clinical trial designs more efficient. Through the retrospective statistical analysis of variation in the design of past trials and the prospective comparisons of clinical trials methods, we can determine which trial procedures truly influence the bias and precision of treatment estimates and where complexity and costs can be reduced. We provide two examples of the retrospective study of clinical trials methods that could change the conduct of future trials. First, an overview of the effect of outcome adjudication on treatment estimates for cardiovascular trials is presented. Second, a prognostic model to detect fraud within multi-centre trials is developed as part of a system of central statistical monitoring. There are many more unanswered questions about efficiencies in clinical trials methodology that need to be examined by statisticians and researchers.

4.1 Background

Researchers have always been interested in studying and improving clinical trials methodology. It is only natural that one who lives by the scientific method may apply this to the process of science itself. Trials today have benefited from many statistical advances including: proper random assignment to treatment groups, intention-to-treat analysis populations, conservative statistical monitoring boundaries, statistical adjustments for multiplicity, proportional hazards models, and many others. To date research has primarily focused on improving trial design and statistical approaches to analyses, but given the changes that are happening, it is the area of trial conduct that now needs our collective thoughts and innovation.

The environment of clinical trials is increasing in complexity and bureaucracy. The current climate for clinical trials is now much more complex, fuelled by

J. Pogue (✉) • P.J. Devereaux • S. Yusuf
Department of Clinical Epidemiology and Biostatistics and Faculty of Health Sciences,
McMaster University, Hamilton, ON, Canada
e-mail: poguej@mcmaster.ca; philipj@mcmaster.ca; yusufs@mcmaster.ca

layers of regulation and a misplaced hypersensitivity and fear of litigation. Take for example, the typical trial informed consent form. This was once a page or two containing an explanation of the purpose of the trial, a summary of the efforts asked of the participant, and a listing any foreseeable risks involved. It is now an equivalent to a corporate contract: over 20 pages is now common, incomprehensible to anyone except lawyers, and full of protection clauses in case of class-action lawsuits, without distinguishing between usual risks seen in clinical practice and any significant risks due to the experimental design or interventions. Thus, due to increasing complexity and bureaucracy today's "informed consent" form no longer fulfils its purpose.

What has happened to the informed consent is mirrored in many parts of clinical trials conduct today where there is a disproportionate focus on minor deviations or inaccuracies in use of inclusion criteria, inappropriate over emphasis on the precision of individual data points and reporting every minor "adverse event" (even those which are part of the natural history of the disease or conditions common in a particular age group). These procedures have gotten in the way of ensuring the precision of the outcomes that matter because so much cost is going into collecting unnecessary data and monitoring aspects of trial conduct that ultimately do not matter.

The danger is that if we cannot determine how to perform well-designed trials in a more efficient manner, we may be left pursuing only expensive small complex trials that have little hope of finding effective treatments for the world's burden of disease. Statisticians and scientists must determine what trial methods, rules, and regulations are necessary, as they really affect trial results and participant safety, and which ones are wasting scientific and monetary resources. Rather than face validity, personal experiences, or legal opinion, what we need now is objective evidence obtained from past and future trials that evaluate whether detailed procedures materially influence trial results and validity.

4.2 Growing Complexity in Modern Trials and its Effect

Prentice [32] pointed out the paradox that the randomized controlled trial, the research design most insulated from confounding, is subject to the most effort and expense to record and control confounders. Compared to cohort studies, trials typically have more complex and detailed inclusion and exclusion criteria, extensive baseline characteristics and follow-up data collection, multiple quality control procedures, standardized outcome monitoring, definitions and reporting, and outcome adjudication. This complexity is ever-present in trials and continues to grow. Getz [18] and Wampler [40] have documented a steady increase in protocol complexities since 1990.

Getz [18] conducted a retrospective analysis of 10,038 phase one to four clinical trials protocols, from pharmaceutical and biotechnology, hosted in the Fast Track System [26]. They estimated a growth in the average total procedures in

trials of 6.5 % per year from 1999 to 2005, with procedures in phase four trial protocols increasing by 9.1 % per year. The total burden of work required of the site investigator by trial protocols increased by 10.5 % per year. The number of eligibility criteria increased three-fold, the median number of reported adverse events grew by a 122 % increase, and the median number of serious adverse events reported per participant in the year 2005 was 12.5 times that of 1990. The average number of case report forms per trial increased from 55 pages to a staggering 180 pages. The length of consent forms has more than doubled, and the work load on REBs has increased substantially yet the number of trials reviewed per month has declined. Performance by sites within trials has been declining within the context of increasing demands. Getz [18] found a 16 % absolute drop in site enrolment between 1999 and 2005, while the rates of retention of participants in trials fell by 21 %.

Others have also documented the growing demands on trial site investigators and participants, the associated increased cost of trials, and an associated decline in site performance. Eisenstein et al. [16] have documented a doubling in the cost of clinical trials over the past decade within the United States from 37 to 64 % of total expenditures of the pharmaceutical industry and the National Institutes of Health from 1994 to 2003. Yet there was a reduction in the number of Food and Drug Administration approvals from 35.5 to 23.3 entities per year over the same period. Clearly this increase in cost has not translated into greater availability of effective disease therapies. Yet the cost of individual trials is substantial, with estimates ranging from 83 to 142 million US dollars for multicenter cardiovascular trials [16]. Trials that require such large investment will by their very nature be limited in number, leading to fewer clinical trials in many areas.

While these “perverse” trends are of great concern, many have suggested that trial methodology can be made more efficient. At the heart of the large simple trial design is the tenant that simple efficient designs will produce the clearest results, and focusing efforts on those methods can influence trial results [12, 43, 44]. Since then others have made further suggestions for increased efficiencies. Thornquist [38] predicted up to a 12 % decrease in total trial budget if non-compliance could be reduced by 50 %. Eisenstein [16] suggested that trials in congestive heart failure and acute coronary syndrome that followed a simplified design could reduce their costs by up to 43 %, without reducing sample size. Simplifications could include: reduced data collection, less use of on-site monitoring, lower site payments, and more efficient site management and data management. Eisenstein et al. [15] suggested that 59 % of the cost of running trials could be saved through reducing planning time for trials, time to recruit the full sample size, reductions in the number of case report forms, smaller numbers of sites, use of electronic data capture systems, and efficient site management practices.

Given that cost and complexity can (and must) be reduced, the question we must now answer is what aspects of trial methodology need to be reduced, improved, increased, or maintained in order to produce reliable, precise, unbiased trial results. We now need researchers experienced in trial methodology to focus their efforts on determining the value of clinical trials practices as a guide to finding needed simplifications and cost reductions. In particular statisticians can assist in this

endeavour by quantifying or measuring the effect of these practices on the treatment effects in terms of bias and precision for both efficacy and safety outcomes. One good source for such methodology research would involve examining the data from past clinical trials, both individually and over multiple datasets to compare and estimate the effect of various clinical trials practices.

All clinical trial practices need to be evaluated in terms of their ability to serve three important functions. First, they may help to minimize the difference between the trial treatment estimate φ and the true effect $g(\theta)$, known as bias in the estimation of the treatment effect:

$$b_{\varphi}(\theta) = E_{\theta}(\varphi(X)) - g(\theta) .$$

While the true effect $g(\theta)$, is never known, it may be expected that a valuable clinical trial practice which reduces bias would bring the estimated treatment effect closer to the true treatment effect. Second, a clinical trial methodology may increase the precision τ of the trial treatment estimate θ or decrease the variance, such that

$$\tau = \frac{1}{\sigma_{\theta}^2} .$$

Lastly, means of performing trial functions efficiently but with less resources, thus reducing the cost of trials, are worthy of study. Means of minimizing bias and increasing precision directly affect trials results, but cost (c) is also indirectly related to precision:

$$\tau \propto \frac{1}{c} .$$

If the cost of enrolling and following an individual subject in a trial is high, trialists may reduce sample size or select less clinically important outcomes in order to make the trial feasible, and this will decrease the precision of the treatment estimate. Having recognised this we now use our units of measurements to examine useful and wasteful procedures in trials. We would like to present two examples where parts of typical trial methodologies were examined to determine their value and suggest possible improvements in efficiency.

4.3 Example 1: Estimating the Effect of Outcome Adjudication

To evaluate the effectiveness of a treatment on a series of protocol-defined outcomes, we require an unbiased collection and validation of these outcomes [17]. One of the key clinical trials methods to ensure this is to have central adjudication of all important outcomes, to determine that they truly meet the protocol definitions.

Outcome adjudication commonly involves a group of experts who examine the supporting documentation for each outcome event and either except it as a valid outcome or reject it from the trial database. The goals of outcome adjudication are clear. Evaluating non-fatal or complex outcomes that may have a subjective component or variability in their ascertainment, in theory could decrease precision due to added “noise” in any trial, regardless of treatment blinding. Adjudication seeks to minimize this potential noise and bias by enforcing standardized outcome definitions through the review of source documents and tests [19, 34, 39]. It should be able to correct any systematic misclassification based on investigators a priori beliefs about the effectiveness of the treatments being compared. Trial credibility is thought to increase if outcome adjudication is used for the trial’s important outcomes [19, 34]. This process is thought to provide a check on the quality and consistency for the trial’s outcomes [14, 19]. This method has appeal to many trialists, as it may increase trial precision by eliminating outcomes that may not be affected by the treatment being studied [14, 19]. It may also help eliminate biased reporting in trials where the therapy is not masked to the participant and/or the site physicians [5, 19, 21, 23]. It is important to recognize, however, that adjudicators typically only evaluate events that were submitted and if biased reporting happens in a trial—whereby there is under-reporting of borderline events in one of the treatment group—this problem is usually not overcome by central adjudication. The process of outcome adjudication does increase the cost of trials [11, 19, 39]. Source documents need to be collected centrally, after redacting all direct participant identifiers, masking and sham masking of treatment information needs to be complete for open trials, translations to the language of the adjudicator may be required for international trials, costs for document shipment, tracking software, and adjudicator remuneration are required.

It remains to be demonstrated if the potential value of outcome adjudication is worth its cost. There have been few systematic efforts to estimate the value of outcome adjudication. Individual trials have occasionally commented that the treatment effect based on investigator-reported outcomes differed from that based on adjudicated outcomes. For example, CHARM-Preserved trial found that candesartan was superior to placebo in reducing the composite of cardiovascular deaths or hospitalization of heart failure with a hazard ratio of 0.85 ($p = 0.028$) based on reported outcomes, but after adjudication this decreased to 0.89 ($p = 0.12$) [45]. The EPIC trial found the opposite pattern that a reported hazard ratio of 0.73 ($p = 0.12$) changed to 0.65 ($p = 0.008$) after the outcomes were adjudicated [36]. Yet other trials, such as the CDP [7] and PARAGON-B [25], have documented consistency between reported and adjudicated trial results. Generally comments about the effect of outcome adjudication on the estimated treatment effect in specific trials are relatively rare and potentially influenced by publication bias.

To address this problem, we conducted a systematic comparison to evaluate and estimate the effect of outcome adjudication within the large cardiovascular trials conducted at the Population Health Research Institute of McMaster University, between 1993 and 2006 [30]. This involved 10 trials with >95,000 trial participants randomized and >9,000 outcomes. It included trials with and without blinding

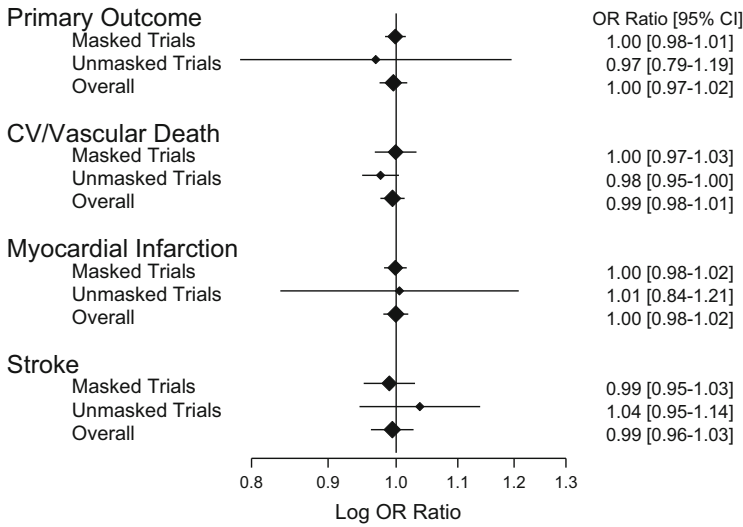


Fig. 4.1 Masked vs. unmasked trials: ratio of odds ratios for adjudicated vs. reported outcomes (Reprinted from [30] with permission)

or masking of treatments. For each trial, we determined the odds ratios for treatment effects using investigator reported events and the treatment odds ratio based on events after adjudication, and pooled these odds ratios across trials with trial as a random effect. The paired difference of the natural logarithm of the odds ratio for adjudicated outcomes minus the natural log odds ratio for reported outcomes was regressed over all trials. Exponentiating this mean difference produced a ratio of odd ratios, where 1.0 indicates no evidence of a treatment difference due to outcome adjudication. This analysis was performed overall and then separately for trials with blinding of treatment group and for trials without blinding. All analyses were weighted by trial size. Figure 4.1 displays the effect of outcome adjudication on the primary outcome for each trial and overall, showing a ratio of odd ratios of 1.00 with 95 % confidence interval 0.97–1.02, implying that we cannot reject the null hypothesis that

$$b_{\varphi}(\theta) = 0 .$$

This estimate was similar for trials with and without blinding [OR ratio = 1.00 (0.98–1.01) and OR ratio = 0.97 (0.79–1.19), respectively]. Similar comparisons were done for individual outcomes included cause specific cardiovascular death, myocardial infarction, and stroke. No significant effect of outcome adjudication was found.

These analyses suggest that the quality monitoring part of systematic and complete outcome adjudication could be eliminated or replaced by a random sampling approach for major cardiovascular mortality and morbidity. Similar

analyses need to be conducted on trials from other coordinating centres and in other research areas (i.e., based on trials with different types of outcomes) so that we may understand when we do and do not need outcome adjudication to minimize bias and maximize precision in trials.

4.4 Example 2: Central Statistical Monitoring as an Alternative to Onsite Monitoring

The gold standard of site monitoring for clinical trials is thought to be frequent on-site visits where all trial data are verified against local source documents. ICH E6 states that, “In general there is a need for on-site monitoring, before, during, and after the trial” [20]. It then goes on to state that central monitoring accompanied by appropriate investigator training and guidance may replace regular on-site monitoring in “exceptional circumstances”. As a result of this guidance document, the use of on-site monitoring is wide spread within industry or clinical research organization trials (84 %), although less commonly used in less well funded academic or government sponsored trials (31 %), based on a survey of 65 research organization that conduct clinical trials in 2009 [27].

On-site monitoring is a costly component of trial methodology, often consuming 20–30 % of the entire cost of a trial, representing tens of millions of dollars for large multi-site trials. Yet there have been surprisingly few evaluations of the effectiveness of on-site monitoring to detect either problems in implementing the trial protocol or possible data fabrication. Published summaries of FDA audits [33] indicate that serious deficiencies are sometime detected (4 % of data audit conducted), but the definition of a serious deficiency is not provided. This means that the reader cannot determine if any of these would have altered trial results. This summary does give examples where data fabrication was detected, but fails to quantify the number of times this misconduct was identified directly by on-site auditors. In contrast to this, others have found that on-site monitoring did not find important problems at sites and did not alter important trial results. The National Institute of Cancer’s on-site monitoring program did not change the agreement rate for treatment failures or the number of protocol deviations [41]. A program of on-site audits started near the end of the GUSTO trial found no errors that would have changed the trial results [17]. The National Surgical Adjuvant Breast and Bowel Project on-site monitoring program found no unknown treatment failures or deaths, and only a very small number of not previously known ineligible participants [9].

In contrast to this dearth of evidence for the effectiveness of on-site monitoring, there have been some limited successes reported with the use of statistical methods and central statistical monitoring to confirm or identify fabricated data. Several authors have used statistical methods to illustrate the implausibility of data sets that were suspected to contain fabricated data. When fraud was suspected in a diet trial submitted for publication to the British Medical Journal, a comparison of these data with that from another diet trial found that their suspicions may have

been warranted. In comparing the intervention to control group within each trial, Al-Marzouki et al. [1] found many more statistically significant differences within the data set thought to be fabricated. Kranke [24] and Carlisle [8] separately used probability models to calculate the chances of observing the group of summary statistics presented in multiple publications ($n = 47$ and $n = 169$) published by a single researcher. Carlisle [8] compared summary binary patient characteristics (e.g., sex or use of antihypertensive medications) to the expected binomial distribution, allowing for a separate population rate per trial across this one researcher's published trials. The discrepancy between these reported and expected distribution was quantified using a Fisher's exact test. For each trial's mean continuous variables (\bar{m}) (e.g., weight or blood pressure) a similar comparison was done using the central limit theorem.

$$\frac{\bar{m} - \mu}{SEM} \times \left(1 + \frac{SD_{SEM}}{\sqrt{SEM}} \right).$$

Here μ is the grand mean over all trials and SEM is the standard error of the mean from each individual trial. They each concluded that these trials collectively resulted in implausible published data. Central statistical monitoring, in various forms, has been used successfully to identify sites within a multi-center trial that have fabricated data. These trials include the AMPIM [3], MRFIT [28], NSABP-06 [9], Second European Stroke Prevention Study [37], COMMIT-1 [10], POISE [13], and other trials. In many of these cases central statistical monitoring identified the problem, while on-site existing monitoring had failed to find the problem.

While the above case studies demonstrate promise for the use of central statistical monitoring in trials, further work in this area is needed. Just as we commonly develop risk models to predict disease in patients, central statistical monitoring could use risk models to identify sites at high-risk for fabricating data, within a multi-centre trial. If a statistical model with sufficient predictive ability could be developed, then their use within central statistical monitoring could replace the function of on-site monitors in fraud detection. We used data from the POISE Trial to retrospectively build a series of prognostic logistic regression models conducted on site-level data to identify the sites that had fabricated their data [29]. Let y take on the value 1 if the site committed fraud and 0 otherwise and suppose there are k independent variables (x_1 to x_k) that predict fraud. Then the probability of fraud having occurred at the j th centre is:

$$p = P\langle y = 1 \mid X = x \rangle ,$$

$$\ln \left(\frac{p_j}{1 - p_j} \right) = \beta_0 = \beta_1 x_1 + \dots + \beta_k x_k .$$

POISE was a multi-centre, multi-national randomized controlled trial testing the effectiveness of a peri-operative beta-blocker in preventing cardiovascular outcome in high-risk patients undergoing non-cardiac surgery. Of the 196 participating

clinical sites from 24 countries, 9 were found to have fabricated data, representing 947 patients out of the total 9,298 randomized within this trial.

For the purpose of building a prognostic model, we used data from all sites that had randomized at least 20 trial patients ($N = 109$ sites). An analytic strategy was followed to develop these prognostic models. First, a wide variety of statistical tests were included since authors have suggested that many types of data and statistics may be useful to identify fabricated data [1–4, 6, 10, 22, 31, 35, 37, 42]. Variables were included from baseline characteristics (binary or continuous), combinations of baseline variables, compliance, site performance, concomitant medications, physical measurements in follow-up, and efficacy and safety outcomes (see Pogue et al. [29] for the complete description). Second, for these models data were summarized at the site level, as opposed to the patient level, since the goal was to identify sites at high-risk for data fabrication. We focused on comparing data across sites and determining how different each site was from the others. We required that these summaries be unit-less, and not dependent on the exact variables collected in the POISE Trial.

These risk models will only be useful for future trials if their prognostic variables may be replaced by the different variables collected in each trial. Making the independent variables unit-less is likely to assist in this goal. Primarily, this involved using probability values (p-values) to quantify how different a site was from all other sites combined for a particular variable. We made no assumptions about direction of effect for these summaries, but instead analysed p-values as continuous possible predictors, rather than using pre-defined cut-offs.

Seven different types of statistical summaries were used. We tested whether each site was different from the rest using a two-by-two frequency comparison for binary variables, such as history of diabetes, and summarized as that site's Pearson chi-square test p-value. We tested how different each site was from the rest for continuous variables (e.g., systolic blood pressure) using two-sample t-tests and calculated a p-value for each site. We compared digit preference for variables such as day of week for randomization for each site versus all others using Pearson's chi-squared test p-values. The variances of continuous variables were compared for each site versus all others using Folded F-test p-values. Distance measures (d_j) were derived for each site for continuous variables indicating how far away one site's data are from the overall mean (\bar{y}) across all centers, standardized by the overall standard deviation (s), using data from the i th trial participant at the j th center. The natural logarithm of distance was used as a possible predictor.

$$d_j = \sum_i \left(\frac{y_{ij} - \bar{y}}{s} \right)^2 .$$

For the comparison of outcomes and compliance, we calculated the probability of observing an outcome rate as extreme as that observed at a site, assuming a Poisson distribution, adjusting for country variation. Instead of testing each center against all the others, we directly calculated each site's cumulative probability distribution (CDF) value from these models. Lastly, for repeated physical measurements over

Table 4.1 Risk scores predicting fabricated data (Reprinted from [29] with permission)

Model terms	Categories	Category	Score
Predictor 1:	SBP over time Intra-class correlations	1	+0
		2	+1
		3	+2
		4	+3
		5	+4
Predictor 2:	DBP Mean Comparison – t-test p-values	1	+0
		2	+1
		3	+2
		4	+3
		5	+4
Predictor 3:	Model 1: SBP digit preference χ^2 p-values	1	+0 (+4)
	Model 2: Surgery: Intrathoracic or intraperitoneal – frequency χ^2 p-value	2	+1 (+3)
	Model 3: Anaesthesia: General – frequency χ^2 p-value	3	+2 (+2)
	Model 4: ACE-I/ARB χ^2 p-value	4	+3 (+1)
	Model 5: Compliance Outcome Probability – CDF	5	+4 (+0)

Note: point reversed for model 5 only and provided in brackets

Categories: ICC and p-values: 1 = ≤ 0.20 , 2 = 0.21–0.40, 3 = 0.41–0.60, 4 = 0.61–0.80, 5 = 0.81+

time, the intra-class correlation coefficient (ICC) itself was used as a unit-less summary for a site's data.

This led to a long list of potential predictors for fabricated data, and we then eliminated redundancy among these using factor analysis with varimax rotation. Out of 52 possible predictors, 18 independent factors were identified and the predictor with the highest loading for each of these factors was selected for inclusion into a series of logistic regression with fraud at each site as the outcome. We used the best subsets of models using the branch and bound algorithm of Furnival and Wilson to find models with the largest score statistic for including different numbers of variables. The final series of models was selected based on no significant increase in the score test for increasing the number of variables in the model. These models were checked for lack of fit using the Hosmer and Lemeshow goodness of fit test. Out of these, the five best predictive models were selected. We then converted these into risk scores using a points system. These are summarized in Table 4.1.

These risk scores were tested in an independent data set in a trial that had on-site monitoring and contained no known data fabrication and produced low-risk score for almost all sites (see Fig. 4.2). These risk scores appear to distinguish well between sites with and without fabricated data, but will require further validation across different types of trials. Where the specific variables used in these score are not collected within a trial, the focus should be on substituting other similar repeated physical measurements or baseline characteristics into these risk scores. The goal is to look for the combination of a site with both greater than normal correlation

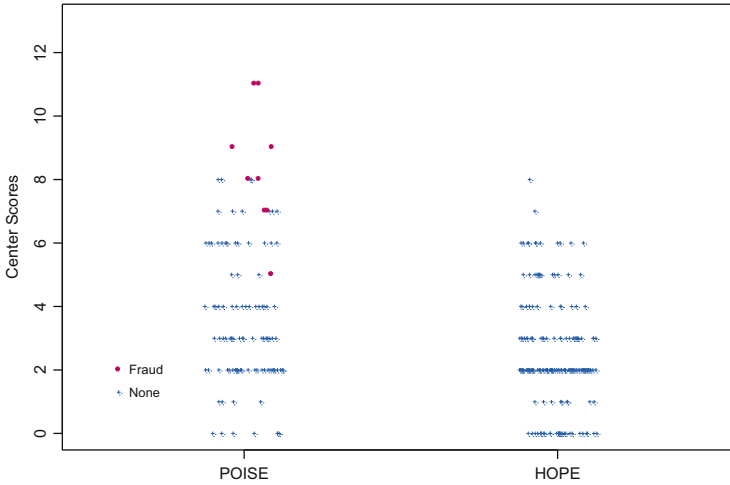


Fig. 4.2 External validation of Model 1 on a trial without fabricated data: a comparison of the distribution of Center Risk Score in POISE (with nine fraudulent centers) and HOPE (no fraudulent centers) (Reprinted from [29] with permission)

over time in physical measurement (high ICCs), and baseline characteristics that look extremely similar to all the other sites (high χ^2 p-values). More research into this area is needed potentially leading to a toolbox of statistical risk scores that can more effectively guide monitoring within trials and lead to greater efficiencies for trials.

4.5 Improving Future Trials

We have illustrated only two investigations into determining what efficient trial conduct should involve. There are many other trial methodologies that need to be studied. It would be useful to estimate the effect of conducting a pilot study prior to launching a full-scale trial, and what are the characteristics of a good pilot study. The effect of complex inclusion/exclusion criteria on speed of recruitment and study power could be estimated. The effect of a run-in period on compliance in the main trial should be studied. These are just a few important unanswered questions that we could examine, using retrospective trial database analyses or overviews of prior trials.

In the future, we may be able to build in tests of differing trial methodology prospectively within given trials. The only way to argue against increasing complexity and bureaucracy is through scientific evidence. We need to separate the elements that matter in conducting a trial that leads to an unbiased, precise answer, from those methodologies that represent a waste of resources. The quality and quantity of future trials may depend on us doing so.

References

1. Al-Marzouki, S., Evans, S., Marshall, T., Roberts, I.: Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *British Medical Journal* **331**, 267–270 (2005)
2. Baigent, C., Harrell, F., Buyse, M., Emberson, J., Altman, D.: Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clinical Trials* **5**, 49–55 (2008)
3. Bailey, K.: Detecting fabrication of data in a multicenter collaborative animal study. *Controlled Clinical Trials* **12**, 741–752 (1991)
4. Benford, F.: The law of anomalous numbers. *Proceedings of the American Philosophical Society* **78**, 551–572 (1938)
5. Boutron, I., Estellat, C., Guittet, L., Dechartres, A., Sackett, D., Hrobjartsson, A., Ravaud, P.: Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: A systematic review. *PLOS Medicine* **3**, 1931–1939 (2006)
6. Buyse, M., George, S., Evans, S., Geller, N., Ranstam, J., Scherrer, B., Lesaffre, E., Murray, G., Edler, L., Hutton, J., Colton, T., Lachenbruch, P., Verma, B., for the ISCB Subcommittee on Fraud: The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine* **18**, 3435–3451 (1999)
7. Canner, P., Krol, W., Forman, S.: External quality control programs. *Controlled Clinical Trials* **4**, 441–466 (1983)
8. Carlisle, J.: The analysis of 169 controlled trials to test data integrity. *Anaesthesia* pp. 1–17 (2012)
9. Christian, M., McCabe, M., Korn, E., Abrams, J., Kaplan, R., Friedman, M.: The national cancer institute audit of the national surgical adjuvant breast and bowel project B-06. *New England Journal of Medicine* **333**, 1469–1474 (1995)
10. COMMIT (CLOpidogrel and Metoprolol in Myocardial Infarction Trial) collaborative group: Addition of clopidogrel to aspirin in 45852 patients with acute myocardial infarction: randomized placebo-controlled trial. *The Lancet* **366**, 1607–1621 (2005)
11. Cook, D., Walter, S., Cook, R., Freitag, A., Guyatt, G., Devitt, H., Meade, M., Griffith, L., Sarabia, A., Fuller, H., Turner, M., Gough, K.: Incidence of and risk factors for ventilator-associated pneumonia in critically ill patients. *Annals of Internal Medicine* **129**, 433–440 (1998)
12. DeMets, D., Califf, R.: Lessons learned from recent cardiovascular clinical trials: Part II. *Circulation* **106**, 880–886 (2002)
13. Devereaux, P., Yang, H., Yusuf, S., Guyatt, G., Leslie, K., Villar, J., Xavier, D., Greenspan, L., Lisheng, L., Xu, S., Malaga, G., Avezum, A., Jacka, M., Choi, P.: Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *The Lancet* **371**, 1839–1847 (2008)
14. Eisenbud, R., Assman, S., Kalish, L., van der Horst, C., Collier, A., for the Viral Activation Transfusion Study (VATS) Group: Differences in difficulty adjudicating clinical events in patients with advanced HIV disease. *Journal of Immune Deficiency Syndrome* **28**, 43–46 (2001)
15. Eisenstein, E., Collins, R., Cracknell, B., Posesta, O., Reid, E., Sandercock, P., Shakhov, Y., Terrin, M., Sellers, M., Califf, R.: Sensible approaches for reducing clinical trials costs. *Clinical Trials* **5**, 75–84 (2008)
16. Eisenstein, E., Lemons II, P., Tardiff, B., Schulman, K.A., Jolly, M., Califf, R.: Reducing the costs of phase III cardiovascular clinical trials. *American Heart Journal* **149**, 482–488 (2005)
17. Friedman, L.M., Furberg, C.D., DeMets, D.L.: *Fundamentals of Clinical Trials*. Springer, New York (2010)
18. Getz, K., Wenger, J., Campo, R., Seguire, E., Kaitin, K.: Assessing the impact of protocol design changes on clinical trial performance. *American Journal of Therapeutics* **15**, 450–457 (2008)

19. Granger, C., Vogel, V., Cummings, S., Held, P., Fiedorek, F., Lawrence, M., Neal, B., Reidies, H., Santarelli, L., Schroyer, R., Stockbridge, N., Zhao, F.: Do we need to adjudicate major clinical events. *Clinical Trials* **5**, 56–60 (2008)
20. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use: *Guidelines for good clinical practice E6. Step 4* pp. 1–53 (1996)
21. Kestle, J., Milner, R., Drake, D.: An assessment of observer bias in the Shunt Design Trial. *Pediatric Neurosurgery* **30**, 57–61 (1999)
22. Knatterud, G., Rockhold, F., George, S., Barton, F., Davis, C., Fairweather, W., Honohan, T., Mowery, R., O’Neill, R.: Guidelines for quality assurance in multicenter trials: A position paper. *Controlled Clinical Trials* **19**, 477–493 (1998)
23. Kradjian, S., Gutheil, J., Baratelle, A., Einstein, S., Kaslo, D.: Development of a charter for an endpoint assessment and adjudication committee. *Drug Information Journal* **39**, 53–61 (2005)
24. Kranke, P., Apfell, C., Roewer, N.: Reported data on granisetron and postoperative nausea and vomiting by Fujii et al. are incredibly nice! *Anesthesia and Analgesia* **90**, 1004–1007 (2000)
25. Mahaffey, K., Roe, M., Dyke, C., Newby, L., Kleiman, N., Connolly, P., Berdan, L., Sparapani, R., Lee, K., Armstrong, P., Topol, E., Califf, R., Harrington, R., for the PARAGON-B Investigators: Misreporting of myocardial infarction end points: results of adjudication by a central clinical events committee in the PARAGON-B trial. *American Heart Journal* **143**, 242–248 (2002)
26. Mathieu, M.P.: Fast track systems. Index of clinical study complexity by year, phases I-IV. In: M.P. Mathieu (ed.) *PAREXEL’s Pharmaceutical R&D Statistical Sourcebook* (2007)
27. Morrison, B., Cochran, C., White, J., Harley, J., Kleppinger, C., Liu, A., Mitchel, J., Nickerson, D., Zacharias, C., Kramer, J., Neaton, J.: Monitoring the quality of conduct of clinical trials: a survey of current practices. *Trials* **8**, 342–349 (2011)
28. Neaton, J., Bartsch, G., Broste, S., Cohen, J., Simon, N., for the MRFIT Research Group: A case of data alteration in the Multiple Risk Factor Intervention Trial (MRFIT). *Controlled Clinical Trials* **12**, 731–740 (1991)
29. Pogue, J., Devereaux, P., Yusuf, S.: Central statistical monitoring: A model to predict fraud in clinical trials. *Clinical Trials* **10**, 225–235 (2013)
30. Pogue, J., Walter, S., Yusuf, S.: Evaluating the benefits of event adjudication of cardiovascular outcomes in large simple RCTs. *Clinical Trials* **6**, 239–251 (2009)
31. Preece, D.: Distribution of final digits in data. *Statistician* **30**, 31–60 (1981)
32. Prentice, R.: Opportunities for enhancing efficiency and reducing cost in large scale disease prevention trials: A statistical perspective. *Statistics in Medicine* **9**, 161–172 (1990)
33. Shapiro, M., Charrow, R.: The role of data audits in detecting scientific misconduct: Results of the FDA program. *Journal of the American Medical Association* **261**, 2505–2511 (1989)
34. Sicurella, J., Roberts, R., Gent, M.: The operation of a central adjudication committee and its effect on the validity of the assessment of treatment benefit. *Controlled Clinical Trials* **11**, 283 (1990)
35. Taylor, R., McEntegart, D., Stillman, E.: Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data. *Drug Information Journal* **36**, 115–125 (2002)
36. The EPIC Investigators: Use of a monoclonal antibody directed against the platelet glycoprotein IIb/IIIa receptor in high-risk coronary angioplasty. *New England Journal of Medicine* **330**, 956–961 (1994)
37. The EsPS2 Group: European stroke prevention study 2. Efficacy and safety data. *Journal of Neurological Sciences* **151**, S1–S77 (1997)
38. Thornquist, M., Urban, N., Tseng, A., Edelstein, C., Lund, B., Omenn, G.: Research cost analysis to aid in decision making in the conduct of a large prevention trial, CARET. *Controlled Clinical Trials* **14**, 325–339 (1993)
39. Walter, S., Cook, D., Guyatt, G., King, D., Troyan, S., for the Canadian Lung Oncology Group: Outcome assessment for clinical trials: how many adjudicators do we need. *Controlled Clinical Trials* **18**, 27–42 (1997)

40. Wampler, S.: Tackling protocol complexity. *Good Clinical Practice Journal* **7**, 6–8 (2000)
41. Weiss, R., Vogelzang, N., Peterson, B., Panasci, L., Carpenter, J., Gavigan, M., Sartell, K., Frei III, E., Mitchel, J., McIntyre, O.: A successful system of scientific data audits for clinical trials: A report from the Cancer and Leukemia Group B. *Journal of the American Medical Association* **270**, 459–464 (1993)
42. White, C.: Suspected research fraud: difficulties of getting the truth. *British Medical Journal* **331**, 281–288 (2005)
43. Yusuf, S., Bosch, J., Devereaux, P., Collins, R., Baigent, C., Granger, C., Califf, R., Temple, R.: Sensible guidelines for the conduct of large randomized trials. *Clinical Trials* **5**, 38–39 (2008)
44. Yusuf, S., Collins, R., Peto, R.: Why do we need some large, simple randomized trials. *Statistics in Medicine* **3**, 409–420 (1984)
45. Yusuf, S., Pfeffer, M., Swedberg, K., Granger, C., Held, P., McMurray, J., Michelson, E., Olofsson, B., Ostergren, J.: Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial. *Lancet* **362**, 777–781 (2003)